8-8-2017

Determination of Dynamical Conservation in Human Cyclophilin Isoforms

Phuoc Jake D. Vu
Georgia State University, pvu4@student.gsu.edu

DETERMINATION OF DYNAMICAL CONSERVATION IN HUMAN CYCLOPHILIN ISOFORMS

by

PHUOC JAKE VU

Under the Direction of Donald Hamelberg, PhD

ABSTRACT

Among the peptidyl prolyl isomerases, the Cyclophilin family of proteins has been linked to various cellular activities such as regulation of homeostasis, mitochondrial permeability, and cell death. Their functionality spans throughout the cell and throughout all cell types as different isoforms. Previous studies done on Cyclophilin A revealed an interesting contact ensemble when bound to a substrate. Because of the similarity of CypA to its homologues, it is believed that they too will exhibit the same contact dynamics. We have defined the dynamics of cyclophilin isoforms through Molecular Dynamics simulations and determined their contact dynamics, characterizing their contact ensembles, and their relative dynamical conservation to each other.

INDEX WORDS: Cyclophilin isoforms, *Cis-trans* isomerization, Peptidyl-prolyl *cis-trans* isomerase, Contact Dynamics, Dynamical conservation, Dynamical Conservation Index

DETERMINATION OF DYNAMICAL CONSERVATION IN HUMAN CYCLOPHILIN ISOFORMS

by

PHUOC JAKE VU

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2017

DETERMINATION OF DYNAMICAL CONSERVATION IN HUMAN CYCLOPHILIN ISOFORMS

by

PHUOC JAKE VU

Committee Chair: Donald Hamelberg

Committee: Gregory Poon

Maged Henary

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

August 2017

DEDICATION

I would like to dedicate this thesis to my parents for always being there to support me throughout this journey.

ACKNOWLEDGEMENTS

First off, I would like to thank my research advisor Dr. Donald Hamelberg for allowing me the opportunity to do research in his lab ever since my junior year of undergraduate studies. As an undergraduate freshman at GSU, I would have never envisioned myself doing research in a lab, let alone sitting here right now writing this thesis. Dr. Hamelberg has exceeded any of my notions of a mentor, having been able to explain the intricacies and the nuances in our field of study in a manner that was always easy to understand. He has been understanding and patient throughout this journey and for that, I am eternally grateful. I would also like to thank my family, friends, and colleagues here at GSU. Special thanks to all my lab mates for helping me whenever I hit wall in my research or whenever I needed to overcome difficult inquiries.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **V**

LIST OF TABLES ..... **VIII**

LIST OF FIGURES ..... **IX**

1 INTRODUCTION **1**

1.1 Cis-Trans Isomerases **1**

1.2 Conservation of Cyclophilin Dynamics **2**

2 MOLECULAR DYNAMICS **4**

2.1 AMBER Force Field **5**

2.2 Integration of Newton's Equation **6**

3 EXPERIMENT **7**

3.1 Human Cyclophilin Isoform Screening **7**

3.2 Preparation of Cyclophilin Systems **8**

3.2.1 Determination of Histidine Protonation States ***8***

3.2.2 Insertion of Substrate into Catalytic Site ***10***

3.3 Molecular Dynamics Simulation of Cyclophilin Systems **11**

3.4 Contact Dynamics **12**

3.5 Dynamical Conservation Protocol **13**

3.5.1 Individual Sequence Alignment of Homologues to Cyclophilin A ***14***

3.5.2 Miscellaneous Individual Sequence Alignment of Homologues ***15***

3.5.3 Cartesian Principal Component Analysis *16*

3.5.4 Dynamical Conservation Plot *16*

3.5.5 Dynamical Conservation Index *17*

4 RESULTS AND DISCUSSION **18**

4.1 Comparison of Cyclophilin A Homologues **18**

4.2 Analysis of Cyclophilin Contact Dynamics **32**

4.2.1 Intra-System *32*

4.2.2 Inter-System *38*

4.3 Cyclophilin Homologue Dynamical Conservation to Cyclophilin A **42**

5 CONCLUSIONS **53**

REFERENCES **54**

APPENDICES **57**

LIST OF TABLES

Table 1-1 Sequence Identities of Cyclophilin Isoforms When Aligned to Each Other..... 3

Table 3-1 Protonation States Used for Each Histidine Found in the Cyclophilin Homologues... 10

Table 3-2 Logistics Used for Alignment of Cyclophilin Homologues to Cyclophilin A..... 15

Table 3-3 Logistics Used for Alignment of Cyclophilin Homologues B, C, D, E to Each Other. 16

Table 4-1 Dynamically Conserved CypA, CypD, and CypE Residue-Residue Contacts of Unbound → Cis. 50

Table 4-2 Dynamically Conserved CypA, CypD, and CypE Residue-Residue Contacts of Cis → TS 51

Table 4-3 Dynamical Conservation Index of Cis Substrate Binding Contacts of Cyclophilin Isoforms. 52

Table 4-4 Dynamical Conservation Index of the Transition of the Cis Substrate Bound State to the TS Substrate Bound State Contacts of Cyclophilin Isoforms. 52

LIST OF FIGURES

Figure 3-1 Clustal Omega Multiple Sequence Alignment of Human Cyclophilin Isoforms A, B, C, D, and E. 9

Figure 3-2 Structural Alignment of Cyclophilins A, B, C, D, and E. 11

Figure 3-3 Sequence Alignment Information of Cyclophilin Isoforms B, C, D, and E Aligned to Cyclophilin A.[49] 15

Figure 4-1 RMSd of CypB systems: Free (A), Cis Substrate Bound (B), V31L mutant (C), and TS Substrate Bound (D). 20

Figure 4-2 RMSd of CypC systems: Free (A), Cis Substrate Bound (B), V31L mutant (C), and TS Substrate Bound (D). 21

Figure 4-3 RMSd of CypD systems: Free (A), Cis Substrate Bound (B), V28L mutant (C), and TS Substrate Bound (D). 22

Figure 4-4 RMSd of CypE systems: Free (A), Cis Substrate Bound (B), V28L mutant (C), and TS Substrate Bound (D). 23

Figure 4-5 RMSd of Cis Substrate Bound CypA. 24

Figure 4-6 RMSF profiles of Unbound Cyclophilin Systems: CypA (black), CypB (red), CypC (green), CypD (blue), and CypE (orange). 25

Figure 4-7 RMSF profiles of Cis Substrate Bound Cyclophilin Systems: CypA (black), CypB (red), CypC (green), CypD (blue), CypE (orange). 26

Figure 4-8 RMSF profiles of TS Substrate Bound Cyclophilin Systems: CypA (black), CypB (red), CypC (green), CypD (blue), CypE (orange). 27

Figure 4-9 Contact Maps of Cyclophilin Cis Substrate Binding Dynamics. 28

Figure 4-10 Contact Maps of the Transition of Cis Substrate Binding to TS Substrate Bound Dynamics. 29

Figure 4-11 Contact Maps Valine to Leucine Mutants versus Wild Type Cyclophilin Dynamics. 30

Figure 4-12 Distance Distribution between Catalytic Site Arginine and Histidine of Wild-Type and Valine to Leucine Mutant Cyclophilin Systems. 31

Figure 4-13 Contact PCA of Cyclophilin A Systems. 33

Figure 4-14 Contact PCA of Cyclophilin B Systems. 34

Figure 4-15 Contact PCA of Cyclophilin C Systems. 35

Figure 4-16 Contact PCA of Cyclophilin D Systems. 36

Figure 4-17 Contact PCA of Cyclophilin E Systems. 37

Figure 4-18 Ranked Contact Index Comparison of Unbound Cyclophilin Isoforms to Unbound Cyclophilin A. 39

Figure 4-19 Contact PCA of Unbound Cyclophilin Homologues to Cyclophilin A. 40

Figure 4-20 Contact Map of the Differences Between the Free State Contact Dynamics of Cyclophilin Isoforms. 41

Figure 4-21 Protein Backbone Cartesian PCA of Unbound Cyclophilin Isoforms CypB (A), CypC (B), CypD (C), CypE (D) compared to those of CypA. 44

Figure 4-22 Plotted Unbound → Cis (left) and Cis → TS (right) Contact Differences of CypB against CypA. 45

Figure 4-23 Plotted Unbound → Cis (left) and Cis → TS (right) Contact Differences of CypC against CypA. 46

Figure 4-24 Plotted Unbound → Cis (left) and Cis → TS (right) Contact Differences of CypD against CypA. 47

Figure 4-25 Plotted Unbound → Cis (left) and Cis → TS (right) Contact Differences of CypE against CypA. 48

Figure 4-26 Plotted Unbound → Cis (left) and Cis → TS (right) Contact Differences of CypD against CypE. 49

1 INTRODUCTION

1.1 *Cis-Trans* Isomerases

Isomerases are a class of enzymes that facilitates the conversion of molecules from one isomer to another.[1] The resulting product has the same molecular formula as the substrate but differs in atomic connectivity or spatial conformation.[2] Isomerases are integral in various biological functions such as glycolysis.[3] Due to the various approaches that isomerases can take for catalysis, they are grouped into 6 different subclasses.[4] The first subclass of isomerase are the racemases and epimerases which inverts the stereochemistry at target chiral carbons.[4,5] The second subclass of isomerase are the *cis-trans* isomerases which catalyze the isomerization of *cis-trans* isomers based on the position of the groups relative to the plane of reference.[4,6] The third subclass of isomerase are the intramolecular oxidoreductases which catalyze the intra-molecular transfer of electrons.[4,7] The fourth subclass of isomerase are the intramolecular transferases which catalyze the intramolecular transfer of functional groups.[4,8] The fifth subclass of isomerase are the intramolecular lyases which catalyze reactions in which a group is eliminated but still remains covalently bonded to the molecule, such as the opening of a ring structure.[4,8] The final subclass of isomerase is not characterized by any specific catalytic approach, rather is a subclass where any uncategorized isomerases are grouped.[4]

The focus of this study is related to the second subclass of isomerase, the *cis-trans* isomerase. Within this subclass, there's a group of enzymes that interconverts the peptide bonds at X-Pro sites from *cis* state to *trans* state and vice versa called peptidylprolyl (prolyl) isomerases PPIase).[9] The confirmation of the X-Prolyl bond can have an effect on protein folding and function based on its current confirmation.[10] Unlike regular peptide bonds, the X-Prolyl bond will not spontaneously adopt its intended conformation therefore peptidylprolyl isomerases are

often used as chaperons for protein folding and regulator of cellular activity.[11] Immunophilins are a group of prolyl isomerase that is composed of two major families, cyclosporine-binding cyclophilins (CyPs) and FK506-binding proteins (FKBPs).[12]

Cyclophilins have been linked to a wide array of cellular activity and isoforms can be found localized throughout the cell. Due to their overall involvement in cellular processes, defects in these proteins can lead to various diseases such as cardiovascular and neurodegenerative diseases.[13,14] Cyclophilin A is generally found in the cytosol of cells and when bound to cyclosporine A, is linked to the suppression of organ rejection via the halt of the production of pro-inflammatory molecules TNF alpha and interleukin 2.[15] Cyclophilin B and C can be found in various tissues types such as breast (B), kidney (C), and bone (C). Studies have shown that the depletion of CypC and CypB can lead to deregulation of the redox homeostasis of the endoplasmic reticulum.[16] Cyclophilin D can be found in the matrix of the mitochondria and the in the nucleus of the cell.[17] When in the matrix of the mitochondria, CypD regulates the permeability of the mitochondrial membrane.[17] When bound to cyclosporine A, CypD's ability to open the pores is suppressed.[17] In the nucleus, CypD plays a role in the regulation of p53.[18] Due to its role in both the mitochondria and nucleus, inhibition of CypD can lead to cell death.[19] Cyclophilin E has an amino-terminal RNA-binding domain and is localized in the nucleus.[20]

1.2 Conservation of Cyclophilin Dynamics

Proteins of the same family will have sequences that are fairly similar and homology of these family members can be determined through their sequence.[21] If two proteins are shown to have a minimum sequence identity threshold of 40-60%, then they should share the same Enzyme Commission number.[22] It would be expected that they should exhibit the same types of catalysis as well. In the case of cyclophilins isoforms, they exhibit sequence identities that fit the

thresholds of 40-60%. This should be reflective in their modes of catalysis and underlying molecular interactions.

Table 1-1 Sequence Identities of Cyclophilin Isoforms When Aligned to Each Other

	A	B	C	D	E
A		64%	63%	75%	68%
B			73%	63%	56%
C				67%	57%
D					69%
E					

Previous Molecular Dynamics (MD) simulations done on CypA revealed an interesting set of dynamics when it is bound to a substrate. When bound to a *cis* substrate, CypA exhibits significant contact ensemble at a site ~15Å away from the catalytic site.[23] We will also take the MD approach to simulate the CypA homologues B, C, D, and E in order to compare their contact dynamics to those of CypA. The contact dynamics of the homologues should mirror that of CypA based on how similar in structure and sequence they are to CypA. This was the first step in the comparison of the dynamics of cyclophilin isoforms. In addition to studies on *cis* substrate binding, transition state (*TS*) substrate simulations of CypA homologues were run and their contact dynamics were defined. We were able to identify certain CypA homologues that exhibited conservation in protein contact dynamics in both the initial binding of the *cis* substrate and the transition of *cis* substrate to *TS* substrate. Key residue-residue contacts in conserved systems were identified as well. We also determined the pairwise contact dynamics conservation of cyclophilin isoforms A, B, C, D, and E.

2 MOLECULAR DYNAMICS

Molecular Dynamics is a computer simulated technique that arose in the late 1970s, which allowed for the study of the underlying dynamics and interactions of micro- and macromolecules.[24] Successful applications of MD simulations have tackled a broad array of issues, such as protein folding, biomolecular association, structure refinement, and allosteric regulation.[25-27] Simulations bridges the macroscopic scales of laboratory settings with the microscopic length and time scales of computational techniques. MD simulations compliments the work of conventional experimentalists, but at the same time, allows us to fill in the gap in information of the discreet molecular interactions that cannot be defined by other means.

MD simulations are achieved through the approximation of the various atomic interactions in a system via quick calculations of a defined potential energy function. The potential energy function is often referred to as a force field, such as AMBER and CHARMM. Each force field are derived slightly differently and results that employs different force fields can vary. The potential of each atom in a system are determined and translated into acceleration and positions over time. Simulations are run over periods of time that are feasible at the time, where the available computer hardware is a limiting factor. They are run long enough so that accurate predictions can be drawn from the bulk properties obtained from the simulations. In the bulk properties lies “hidden” details on the underlying interactions of the simulated system. This allows for the interpretation biological systems that are not accessible by purely experimental methods and measurements and opens doors to opportunities for comparative and collaborative studies.

2.1 AMBER Force Field

MD simulations employs the usage of a potential energy function of molecular parameters to describe the inter- and intra- molecular forces between atoms in a simulated system. It consists of approximations that forms a force field consisting of several terms that provide the contribution of bonded and non-bonded interactions. This is performed in order to lessen the computational stress imposed on the computer and to maximize the speed of calculations. The bonded interactions describe the ability of bonds to stretch, bend, and torsion, which are defined by the bond length, angle, and dihedral parameters. The non-bonded interactions describe the electrostatics and van der Waals interactions, which are defined by the Columbic potential and Lennard-Jones potential. The AMBER force field was used to carry out all the simulations in this study.[28]

$$U(r_1,\cdots,r_N)=\sum_{bonds}\frac{K_r}{2}(l_i-l_{i0})^2+\sum_{angles}\frac{K_\theta}{2}(\theta_i-\theta_{i0})^2+\sum_{torsions}\frac{K_\phi}{2}[1+\cos(n\phi_i-\gamma_i)]+\sum_{atoms\ pairs}4_{\varepsilon ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}-\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]+\sum_{atom\ pairs}\frac{q_iq_j}{\varepsilon_l r_{ij}} \qquad \textbf{(2.1)}$$

The potential energy function presented in equation 1.1 includes the intra-molecular interactions, defined in the first three terms, and the inter-molecular interactions, defined in the last two terms. The parameters K_r, K_θ, and K_ϕ accounts for the bond, angle, and dihedral force constants. $l_i - l_{i0}$ accounts for the changes in bond length and equilibrium bond length. $\theta_i - \theta_{i0}$ accounts for the change in equilibrium angle between three atoms. In the third term, $n\phi_i - \gamma_i$, n represents the multiplicity, ϕ_i is the dihedral angle, and γ_i is the phase shift. The fourth term, the first of the two inter-molecular terms, uses the Lennard-Jones potential to account for the repulsive, $\left(\frac{1}{r_{ij}}\right)^{12}$, and attractive, $\left(\frac{1}{r_{ij}}\right)^{6}$, forces between two particles. r_{ij} represents the distance

between two particles i and j, εij represents the Lennard-Jones well depth, and σ_{ij} represents the distance between two particles i and j at the minimum potential obtained from experimental data or quantum calculations. The fifth term, uses the Coulomb potential to describe the electrostatic interaction. $q_i q_j$ represents the point charges of q_i and q_j, r_{ij} distance between two atoms i and j, and ε_l represents the effective dielectric constant.

2.2 Integration of Newton's Equation

Once the initial coordinates and velocities are determined for all atoms in a system, a trajectory was generated from the integration of a set of classical Newton's equations of motion.

$$F_i = m_i \frac{d^2 r_i(t)}{dt^2} \quad \textbf{(2.2)}$$

In equation 2.3, F_i is the force acting on an atom, m_i is the mass of the atom, and $r_i(t)$ is the position of the atom at a specific time. The acceleration is expressed as the second derivative of the position of the atom relative to time t. Force can also be expressed as the gradient of potential energy:

$$F_i = -\nabla_{ri} U(r_1, \cdots, r_N) = -\left(\frac{\partial U}{\partial x_i}, \frac{\partial U}{\partial y_i}, \frac{\partial U}{\partial z_i}\right) \quad \textbf{(2.3)}$$

The correlation between the derivative of potential energy and the changes in position as a function of time is described through the combination of equations 2.3 and 2.4.

$$-\left(\frac{\partial U}{\partial x_i}, \frac{\partial U}{\partial y_i}, \frac{\partial U}{\partial z_i}\right) = m_i \frac{d^2 r_i(t)}{dt^2} \quad \textbf{(2.4)}$$

The accelerations of all atoms are determined through the forces acting on the atoms and subsequently, new positions and velocities of every atom in a system are derived over a period of time called a timestep.[29] The new positions and velocities are used to recalculate the forces in iterative steps over an interval of time, where the positions, velocities, and force are saved to

generate a trajectory of the system. The timestep that is typically used is on the femtosecond timescale due to the vibrational frequencies of heavy atom hydrogen bonds in a molecular system corresponds to ~10 fs. Simulations that employs shorter timesteps are more computationally expensive. This is overcome through the usage of the SHAKE algorithm, which allows for the usage of longer timesteps, i.e. 2fs, in MD simulations.[30] The SHAKE algorithm constrains all hydrogen related bonds in order to remove the highest vibrational frequency in a system and is based on the Verlet algorithm derived from the two Taylor expressions.[31]

$$r_i(t + \Delta t) \cong r_i(t) + (v(t)\Delta t) + \frac{F_i(t)}{2m_i}\Delta t^2 \quad \textbf{(2.5)}$$

$$r_i(t - \Delta t) \cong r_i(t) - (v(t)\Delta t) + \frac{F_i(t)}{2m_i}\Delta t^2 \quad \textbf{(2.6)}$$

In these expressions, $r_i(t)$ refers to the position at a certain timestep, $r_i(t + \Delta t)$ and $r_i(t - \Delta t)$ refers to the position at the next and previous timestep respectively. The positions are calculated with no use of explicit velocities. The expressions can be combined to obtain:

$$r_i(t + \Delta t) \cong 2r_i(t) - r_i(t - \Delta t) + \frac{F_i(t)}{m_i}\Delta t^2 \quad \textbf{(2.7)}$$

The combined expression takes into account the positions and accelerations at time t and the positions from time $r_i(t - \Delta t)$ to calculate new positions at time $r_i(t + \Delta t)$.

3 EXPERIMENT

3.1 Human Cyclophilin Isoform Screening

Protein Data Bank (PDB) IDs of cyclophilin B, C, D, E, and F (CypB, CypC, CypD, CypE, and CypF) were compiled and then screened for sequence fidelity and high resolution. Due to recent classification changes, cyclophilin D and F are currently the same protein based on sequence alone. For the purpose of this and future work, cyclophilin D/F will be referred to as

cyclophilin D (CypD). Based on the compiled list, the following cyclophilins were chosen for the next steps of this experiment: (1) CypB = 3ICH (2) CypC = 2ESL (3) CypD = 4O8H (4) CypE = 3UCH. Those protein structures were then downloaded and 2 copies of each files were saved.

3.2 Preparation of Cyclophilin Systems

Excess small substrates were and artifacts of crystallization were removed in both copies of each systems. In the PDB for CypB, a double lysine KK tag had to be removed from the beginning of the sequence. In the PDB for CypC, an arginine tag was removed from the beginning of the sequence. In the PDB of CypE, all the starting residues up to the first serine in the sequence was removed. In addition to the small substrates and artifacts, crystal waters were removed from one copy of each system. Three systems were prepared from the copy with crystal waters intact: (1) unbound WT protein (2) unbound V31L (CypB and CypC)/V28L (CypD and CypE) mutant protein (3) *cis* substrate bound WT protein. An additional *TS* substrate bound WT protein system was generated based off the equilibrated state of the *cis* substrate bound WT protein system.

3.2.1 Determination of Histidine Protonation States

The web-based sequence alignment tool Clustal Omega was used to determine any conserved residues, such as histidine, and the protonation state of said residues.[32-34] According to Figure 2.1, the 4 histidine found in CypA were conserved in the isoforms, with the exception of H70 of CypA. Those conserved histidine were altered to match the same histidine protonation state used in the previous studies of CypA. The altered files were the ones that still had the crystal waters intact. In order to determine the protonation state of the non-conserved histidine, the web-based pKa determination tool H++ version 3.2 was used.[35-38] The stripped PDB files of each cyclophilins were inputted into the tool under the conditions of pH 7.0, salt concentration of

0.15, external dielectric of 80, and internal dielectric of 10. Once the tool completed its calculations, the protonated atom PDB files were downloaded and viewed in VMD and compared to the original file with the crystal waters to determine if the protonation states were valid enough to proceed. The determined protonation states of the histidine for each system are presented in Table 2.1. The non-conserved histidine were altered in the same file as the conserved histidine to match the agreed upon protonation state.

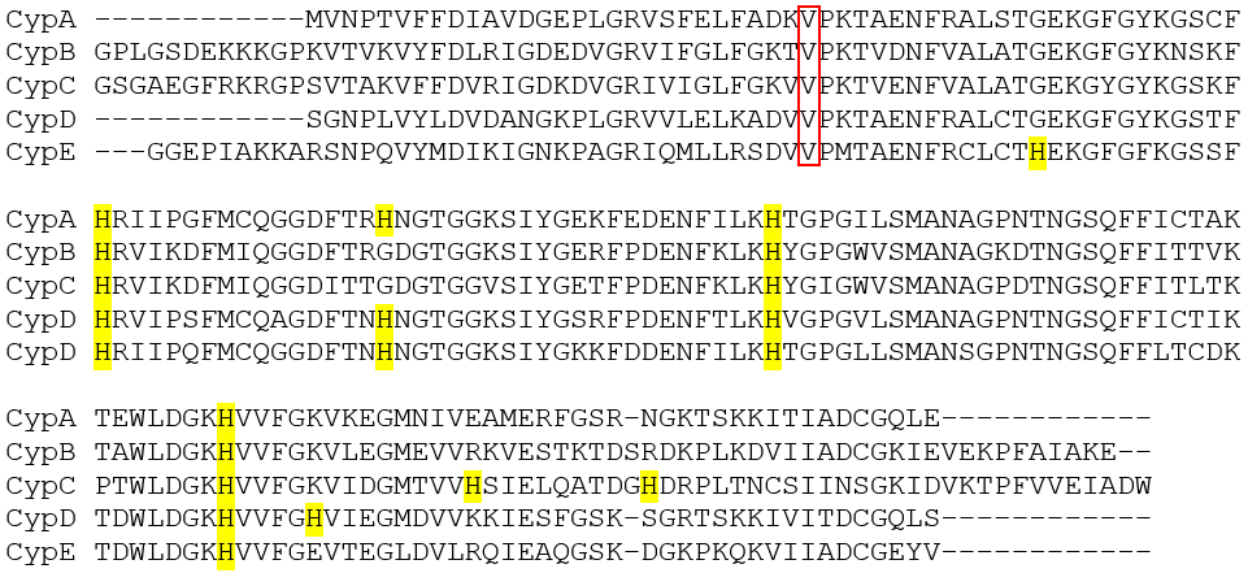


Figure 3-1 Clustal Omega Multiple Sequence Alignment of Human Cyclophilin Isoforms A, B, C, D, and E.

Table 3-1 Protonation States Used for Each Histidine Found in the Cyclophilin Homologues.

A		B (+4)		C (+31)		D (+1)		E	
-		-		-		-		50	E
54	E	58	E	57	E	53	E	62	E
70	P	-		-		69	P	78	P
92	D	96	D	95	D	91	D	100	D
126	P	130	P	129	P	125	P	134	P
-		-		-		130	E	-	
-		-		141	P	-		-	
-		-		153	D	-		-	

3.2.2 Insertion of Substrate into Catalytic Site

The bound states were created through the overlapping of the unbound protein of interest with the pre-prepared *cis* substrate, **Gly-Ser-Phe-Gly-Pro-Asp-Leu-Arg-Ala-Gly-Asp,** bound state of CypA. The two files were loaded into PyMOL and the unbound protein was structure aligned to the bound CypA.[39] In order to generate a new coordinate file with the substrate and the protein, the atoms of the CypA protein were deleted. Once protein portion of the CypA file were gone, the coordinates of the new system was saved as a PDB. The new PDB files were examined for the presence of the TER marker that indicates the separation of the protein residues and the substrate residues. The PDB files were inputted in xLeap to confirm the fidelity of the insertion.

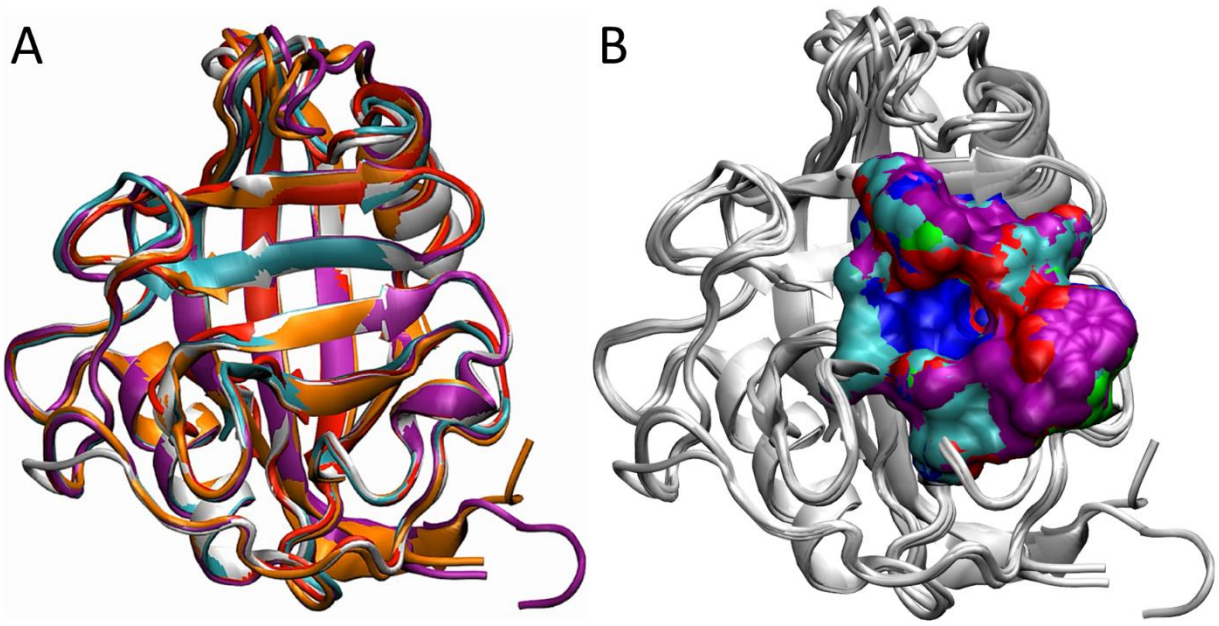


Figure 3-2 Structural Alignment of Cyclophilins A, B, C, D, and E.
(A) The initial structures of CypA (white), CypB (orange), CypC (purple), CypD (red), and CypE (cyan) were aligned in PyMOL. (B) The surface of the active sites of CypA (green), CypB (blue), CypC (Purple), CypD (red), and CypE (cyan) were overlaid.

3.3 Molecular Dynamics Simulation of Cyclophilin Systems

Simulations were performed using the AMBER14 package of programs along with AMBER ff14SB force field[28,40] and the re-optimized dihedral parameters[41] for the backbone ω-bond angles. Each cyclophilin systems were solvated with TIP3P water in a periodic octahedron box with a 10 Å spacing distance between the protein and edges of the box. Additional sodium or chloride ions were added to neutralize the net charge of the protein. Each system was put through 3 minimization steps followed 1 equilibration step. The *TS* system was generated based on the equilibrated state of the *cis* substrate bound systems. The umbrella sampling method was used to transition the substrate from the *cis* state to the *TS* state where a harmonic restraint was applied to the ω bond angle of the proline. The ω bond angle was rotated

by ten degrees from the *cis* ω=0° to the *TS* ω=90°.[42] The final ω=90° sampled state was used as the equilibrated *TS* substrate bound system for each homologue.

Long-range electrostatics were evaluated with the particle mesh Ewald method. A cutoff of 9 Å was used for the evaluation of the short-range nonbonded interactions. The SHAKE algorithm was employed to constrain the bonds between heavy atoms and hydrogens. A 2 fs timestep was used in the integration of Newton's equations of motion. Trajectories were saved every 1 ps. Simulations were run until a stable 2 μs of stable simulated time was reached. Root mean square deviation (RMSD) calculations were performed to confirm the overall stability of the protein in each system. Root mean square fluctuation (RMSF) calculations on the protein of each systems were evaluated. The output was then processed to align the residues of the homologues to that of CypA in accordance to Figure 2.2 and Table 2.2. The distance of the catalytic site of the unbound and mutant systems were measured based on the equivalent R55Cζ and H126Cα in each system. The distances were compared to that of CypA of Doshi et al.

3.4 Contact Dynamics

Individual PDBs were generated for every frame in the 2.0 microseconds of stable simulated time in each system; a total of 2 million frames are present in each system. The PDBs are then analyzed to determine every possible combination of residue to residue contacts. If a heavy atom of a residue is within 4.5 Å of another atom of another residue; that interaction will be recorded as a contact.[43-47] Once all 2 million frames are analyzed, a single file is generated of every contact combinations and number of occurrence called a contact statistics file. The contact frequency, f_i, was calculated based on $f_i = t_i/T$, where t_i is the number of occurrence and T is

the number of frames analyzed. An average of the contact frequencies were made whenever we compared a group of systems.

Contact maps were generated based on the difference between 2 systems. The difference can represent the formation of bonds or the breaking of bonds. The difference files were screened against the average statistics file where if the average frequency was $0.1 < f < 0.9$, the corresponding difference were retained while $f < 0.1$ and $0.9 < f$ were omitted. The values that remained was used to generate a file that represents the possible contact formation and breaking when the protein shifts from one state to another state. If the difference probability was greater than or equal to 10%, they were represented as cylinders that connects the Cα of two corresponding residues.

Meaningful contacts files were generated based on if $0.1 < f < 0.9$ in the average contacts. The meaningful contacts files were used to generate contact trajectories that contained a N X C binary matrix of 0s and 1s, where N is the total number of frames analyzed and C is the number of meaningful contacts. Principle component analysis was then performed on the combined contact trajectories of interest. MATLAB was used to calculate and diagonalize the covariance matrix as well as the projection of the data onto the first two principal eigenvectors of the contact space.[48]

3.5 Dynamical Conservation Protocol

Comparison of the cyclophilin isoforms required various steps that identifies consensus residue lengths in addition to verification of fidelity of the process. Comparisons were limited to only two isoforms at a time. Once consensus sequences for each observable pair were determined, the contact statistics files for each pair were processed accordingly. The line counts were an indicator used to determine the fidelity of the file processing; the line counts should be

the same for every system in each pair. Once a set of logic was determined for each pair, a series of analysis were performed to qualitatively compare the isoforms.

3.5.1 Individual Sequence Alignment of Homologues to Cyclophilin A

The FASTA sequence of the homologues were aligned to that of CypA with the online alignment tool BLASTP 2.6.1+.[49,50] The alignment results are presented in Figure 2.2. The alignment numbers were reconfigured based on the residue numbers of the simulated systems. A set of logic was determined based on the reconfigured alignments, shown in Table 2.2. The file pairs of interest were run through preliminary processing in which their line counts were compared for fidelity of the alignment. If the line counts do not equal, then the alignment logic was tweaked until a consensus line count was achieved. This method was used to prepare the files for the comparison of the unbound state transition to the *cis* substrate bound state (unbound → *cis*) and *cis* substrate bound state transition to the *TS* substrate bound state (*cis* → *TS*) differences and the unbound internal motions between: (1) CypA and CypB (2) CypA and CypC (3) CypA and CypD (4) CypA and CypE.

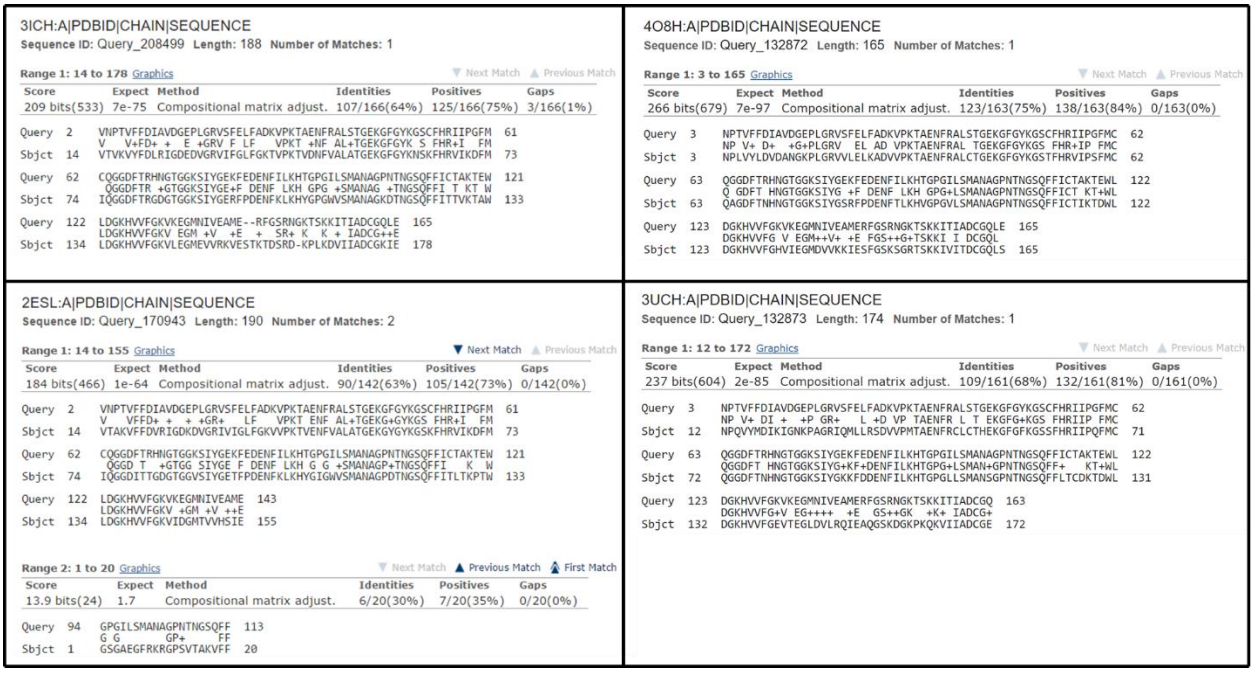


Figure 3-3 Sequence Alignment Information of Cyclophilin Isoforms B, C, D, and E Aligned to Cyclophilin A.[49]

Table 3-2 Logistics Used for Alignment of Cyclophilin Homologues to Cyclophilin A.

A-B		A-C		A-D		A-E	
A	**B**	**A**	**C**	**A**	**D**	**A**	**E**
x>1 **x<166** **x!=150**	**x>3** **x<169** **x!=146** **x!=147**	**x>1** **x<166**	**x>3** **x<169** **x!=150**	**x>2** **x<166**	**x>1** **x<165**	**x>2** **x<164**	**x>1** **x<163**

3.5.2 Miscellaneous Individual Sequence Alignment of Homologues

The FASTA sequence of the homologues were aligned in various combinations with the online alignment tool BLAST. The alignment numbers were reconfigured based on the residue numbers of the simulated systems. A set of logic was determined based on the reconfigured alignments, shown in Table 3.3. The same fidelity checking methods were used as in section 2.5.1. This method was used to prepare the files for the comparison of the unbound → *Cis* and

Cis → *TS* differences between: (1) CypB and CypC (2) CypB and CypD (3) CypB and CypE (4) CypC and CypD (5) CypC and CypE (6) CypE and CypD.

Table 3-3 Logistics Used for Alignment of Cyclophilin Homologues B, C, D, E to Each Other.

B-C		B-D		B-E		C-D		C-E		E-D	
B	C	B	D	B	E	C	D	C	E	E	D
x>0 x<176	x>0 x<176	x>7 x<169 x!=152	x>4 x<165	x>0 x<163 x!=159	x>3 x<167 x!=147 x!=148	x>4 x<143	x>7 x<146	x>1 x<147	x>4 x<150	x>1 x<163	x>1 x<163

3.5.3 Cartesian Principal Component Analysis

Cartesian PCA was run for the unbound homologues paired with the unbound CypA. The trajectories of each pairs were recompiled so that only the coordinates of the aligned residues and their backbone heavy atoms remained. In addition, a new parameter and topology file was generated based on the same method used to recompile the trajectories. CPPTRAJ of the AMBER14 suite of programs was used for the Cartesian PCA calculation. The correlated internal motion of each system was represented as a covariance matrix or Cartesian coordinates. The covariance matrix was then diagonalized to obtain the first 3 principal eigenvalues. The first two principal eigenvalues were then projected.

3.5.4 Dynamical Conservation Plot

The aligned contact statistics files of the unbound protein state, *cis* substrate bound state, and *TS* substrate bound state for each comparison pairs were gathered into a single location. The difference in contact frequencies were determined for the following states: (1) unbound → *cis* (2)

cis → *TS*. Those differences were further processed for any contact pairs that did not meet the following cut off: if the difference of CypA is not greater than ±.1 or if the difference of CypD/E is not greater than ±.1 ((sqrt($3*$3)>.1) || (sqrt($6*$6)>.1)). The remaining pair wise values were plotted against each other. The plot was divided into 4 quadrants with the x-axis and y-axis as separation markers. Shared conserved residue-residue contacts were identified according to defined dynamical conservation in section 3.5.5. The residue-residue contacts of the combination pairs of CypA, CypD, and CypE were extracted from the dynamical conservation plots and were screened for any shared dynamically conserved contact pairs.

3.5.5 Dynamical Conservation Index

Information from the dynamical conservation plot analysis was used as a base for the determination of the dynamical conservation index of each isoform pairs. Residue-Residue contact pairs were assigned numeric values based on their dynamical conservation, i.e. where they lie on the plots. Pairs that were located in the top right and bottom left quadrants were considered dynamically conserved pairs and were assigned the number 1. Pairs that were located in the top left and bottom right quadrants were considered dynamically non-conserved pairs and were assigned the number -1. Pairs that had one coordinate value that lied on an axis were considered dynamically non-conserved but were assigned the number 0. Those assigned values were averaged to determine the dynamical conservation index number.

4 RESULTS AND DISCUSSION

4.1 Comparison of Cyclophilin A Homologues

Preliminary examinations for stability and comparisons of CypA homologue systems results to those of Doshi et al. were performed to verify the fidelity of our results.[23] The RMSD of each homologue systems suggests that the overall stability of the system was reached. The RMSD results also suggests some insight into the dynamics of each homologues. CypC had the widest range in deviation which suggests that it was the most dynamic of the homologues. Contrary to CypC, CypD's deviations were fairly limited which suggests that it was the least dynamic. In addition to the homologues, CypA was re-ran with the longer *cis* substrate. The RMSD of the *cis* substrate bound CypA suggests that the protein was stable. The backbone RMSF of each homologue and system were compared to that of CypA. This was done as another metric of fidelity of the residue alignments of the homologues to CypA. The shared fluctuations of certain residues presented in Figures 4.6-4.8 suggests that the residue alignments were performed correctly.

The comparison of the maps of the *cis* substrate bound CypA of Doshi et al. and our *cis* substrate bound CypA revealed that the unique cluster of contacts 80~ Å away from the catalytic site was still present. It was expected that this unique interaction is conserved in CypA homologues. As presented in Figure 9, we do not observe a conservation of the contact ensemble throughout the homologues. According to the contact maps, only CypE had similar contact dynamics to that of CypA, while CypD was highly stable in its contacts. In addition, CypB and CypC were similar in their contacts. Based on the phylogenetic tree and the sequence identities, we expected that CypD would be the most similar to CypA but it turned out to be the least similar. CypB and CypC were expected to be similar to one another and that was reflected in

their contact dynamics. Due to the lack of conservational dynamics with the *cis* substrate binding, lead us to pursue the dynamics present during the transition from the *cis* substrate state to the *TS* substrate state. The contact maps presented in Figure 10 revealed that the contacts of the transition of the *cis* substrate to *TS* substrate were fairly similar across the homologues. This would suggest that it is the not the initial binding dynamics that are conserved, but the dynamics of the catalysis that are conserved.

The results of the leucine mutants presented in Figures 4.9-4.10 further suggests the high similarity of CypE to CypA. The contact maps of the homologues were similar to that of the *cis* substrate binging contacts. The distance distribution between the guanidinium carbon of the side chain of Arg-(CypB/CypC(57), CypD/CypE(54)) and backbone Cα atom of His-(CypB/CypC(128), CypD/CypE(125)) revealed that CypB and CypC only had 2 distinct populations while CypD and CypE had 3 distinct populations. This suggests that CypD and CypE are similar to the CypA distances of Doshi et al. in terms of number of populations. However, they do not necessarily shared the same shifts in populations.

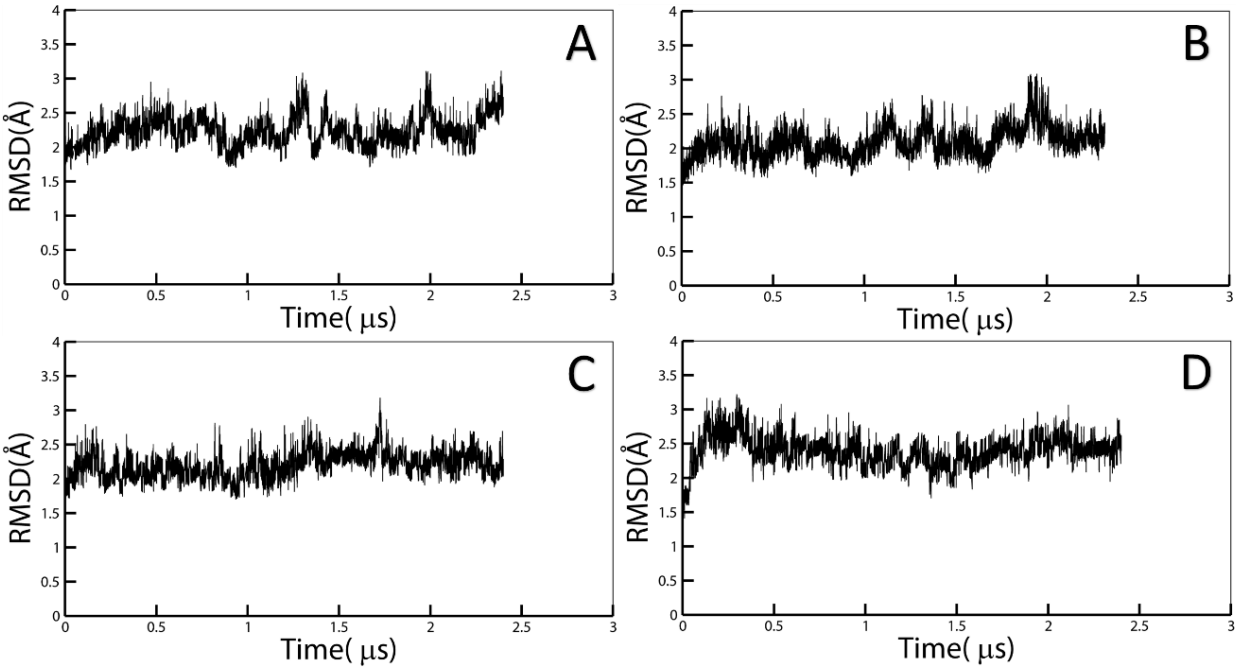


Figure 4-1 RMSd of CypB systems: Free (A), *Cis* Substrate Bound (B), V31L mutant (C), and *TS* Substrate Bound (D).
The free and V31L systems were imaged once for the protein. The *cis* and *TS* substrate bound systems were imaged twice, once for the substrate and once for the entire system (substrate+protein). RMSd was only done on the protein.

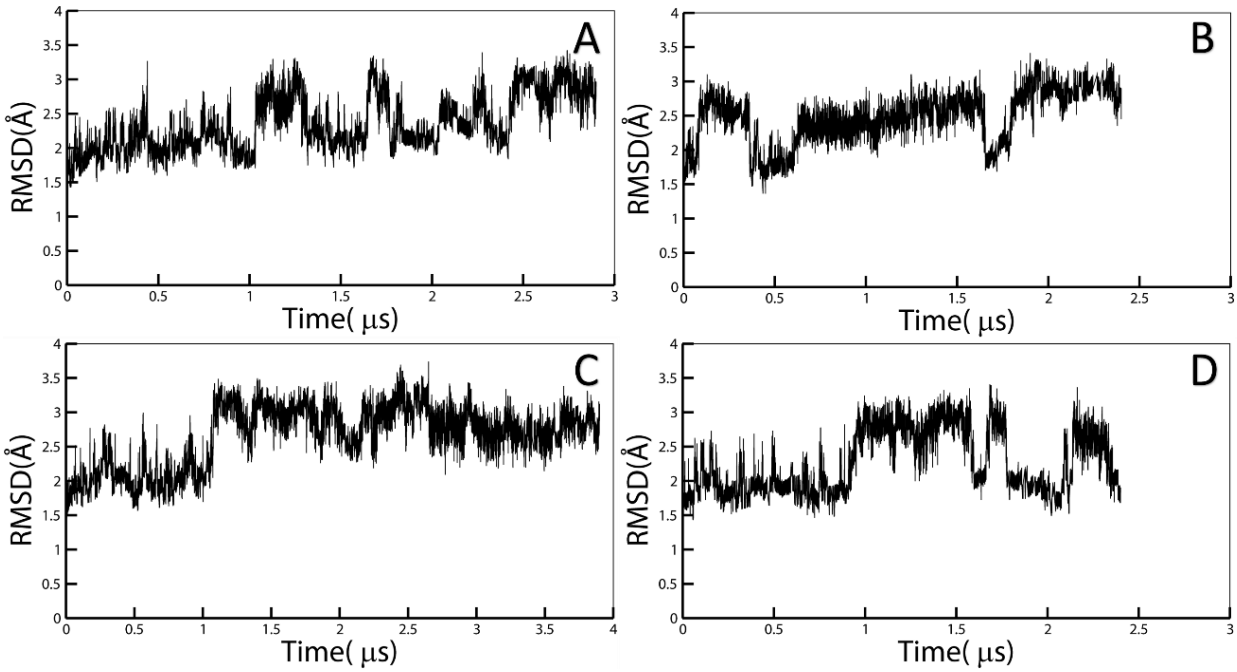


Figure 4-2 RMSd of CypC systems: Free (A), *Cis* Substrate Bound (B), V31L mutant (C), and *TS* Substrate Bound (D).
The free and V31L systems were imaged once for the protein. The *cis* and *TS* substrate bound systems were imaged twice, once for the substrate and once for the entire system (substrate+protein). RMSd was only done on the protein.

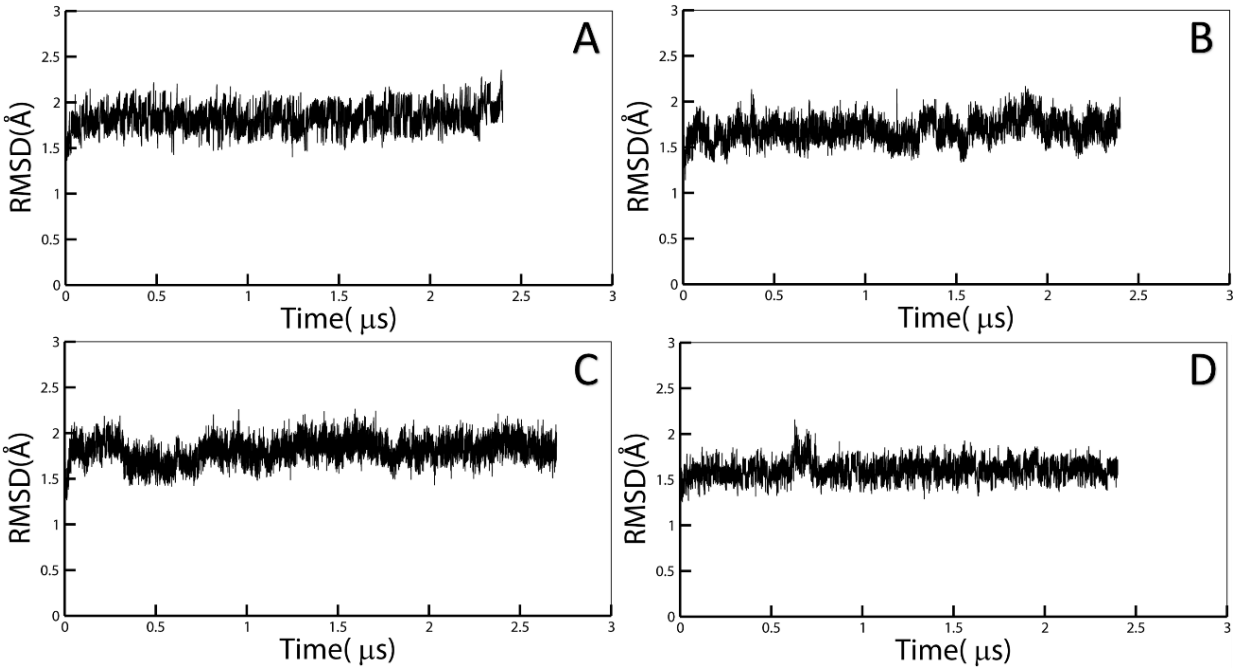


Figure 4-3 RMSd of CypD systems: Free (A), *Cis* Substrate Bound (B), V28L mutant (C), and *TS* Substrate Bound (D).

The free and V28L systems were imaged once for the protein. The *cis* and *TS* substrate bound systems were imaged twice, once for the substrate and once for the entire system (substrate+protein). RMSd was only done on the protein.

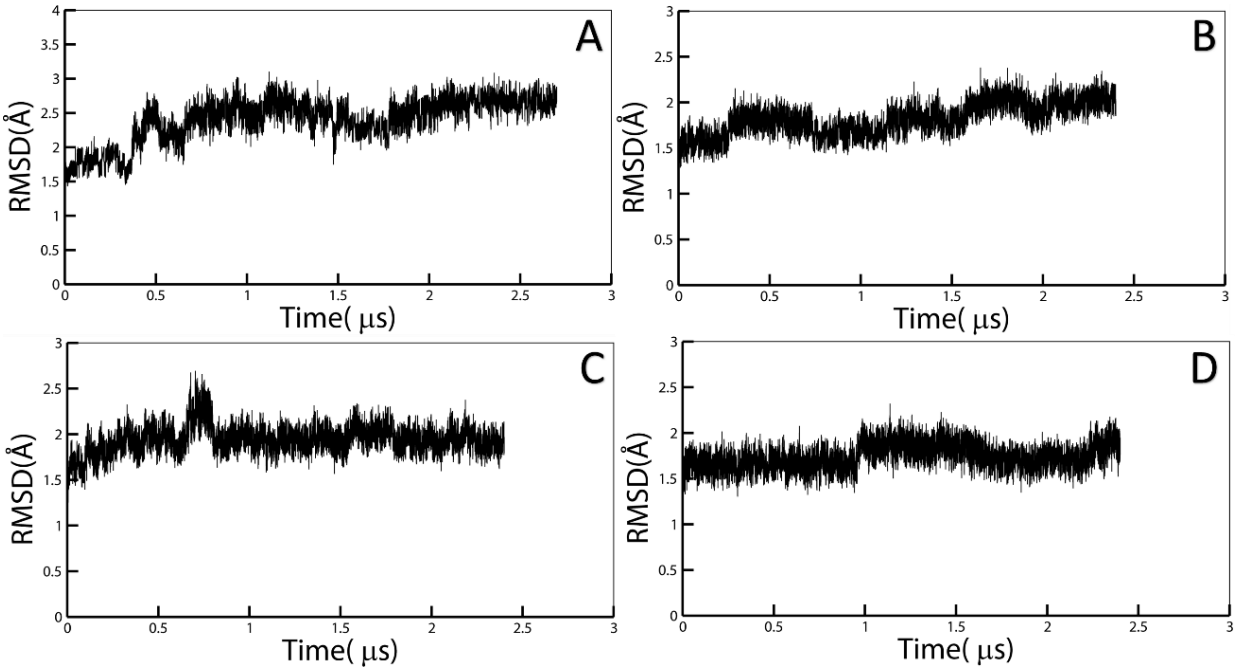


Figure 4-4 RMSd of CypE systems: Free (A), *Cis* Substrate Bound (B), V28L mutant (C), and *TS* Substrate Bound (D).
The free and V28L systems were imaged once for the protein. The *cis* and *TS* substrate bound systems were imaged twice, once for the substrate and once for the entire system (substrate+protein). RMSd was only done on the protein.

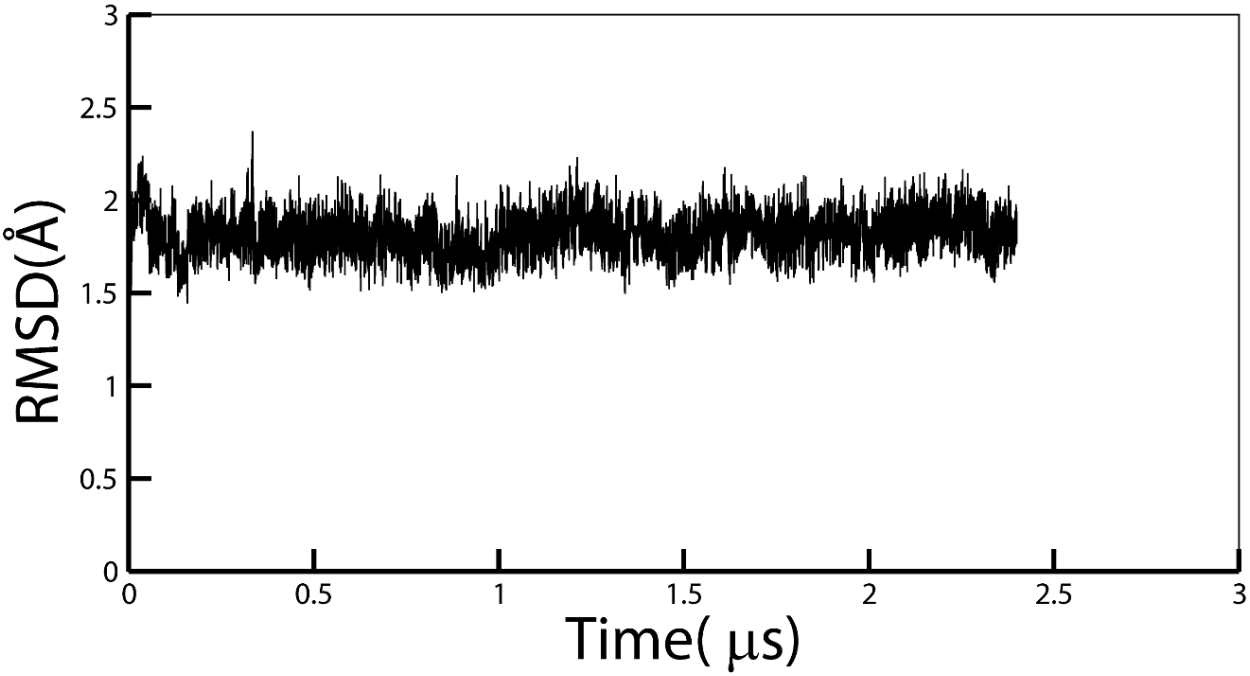


Figure 4-5 RMSd of *Cis* Substrate Bound CypA.
The system was imaged twice, once for the substrate and once for the entire system (substrate+protein). RMSd was only done on the protein.

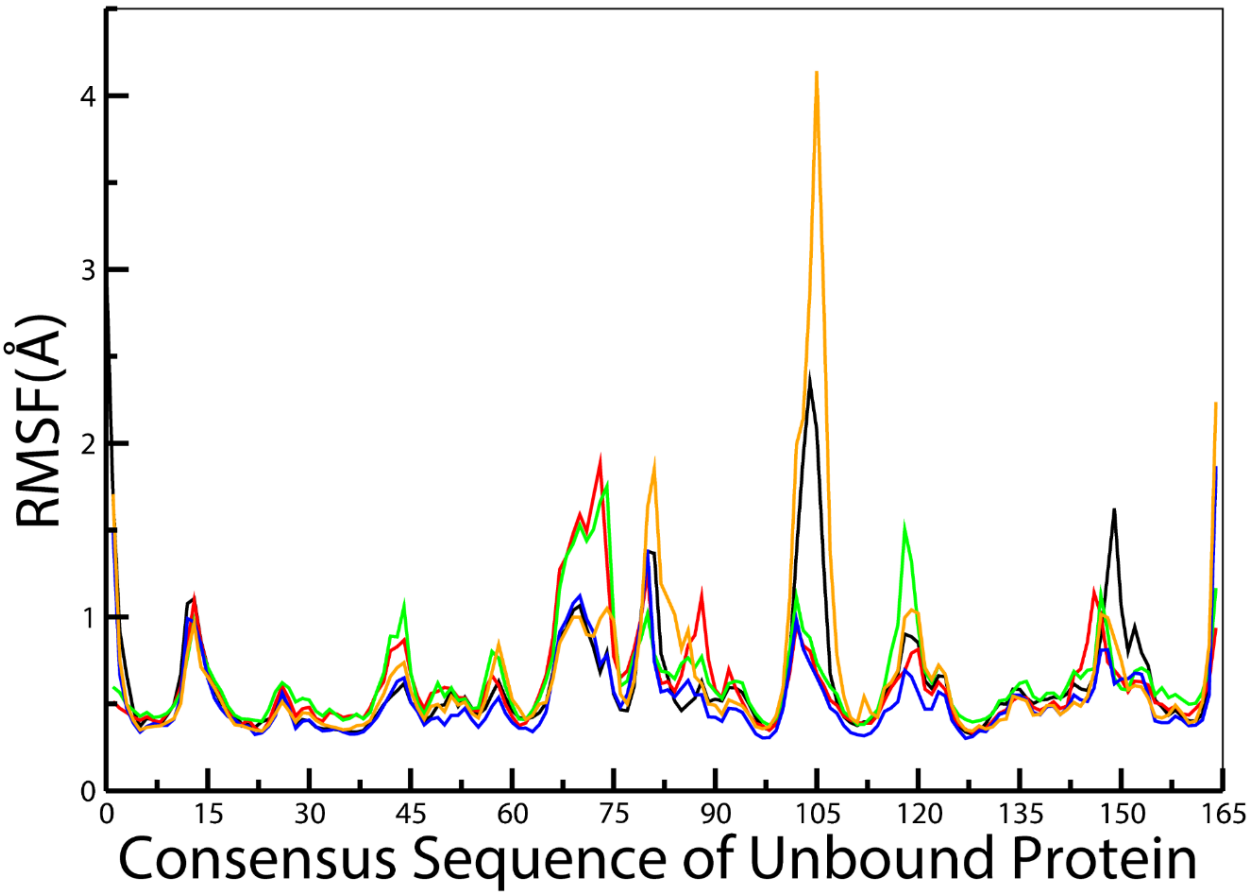


Figure 4-6 RMSF profiles of Unbound Cyclophilin Systems: CypA (black), CypB (red), CypC (green), CypD (blue), and CypE (orange).
The RMSF for each system were initially ran on their entire protein sequence length. The output files were processed to align the residues of each system to that of CypA, in accordance to the individual sequence alignments.

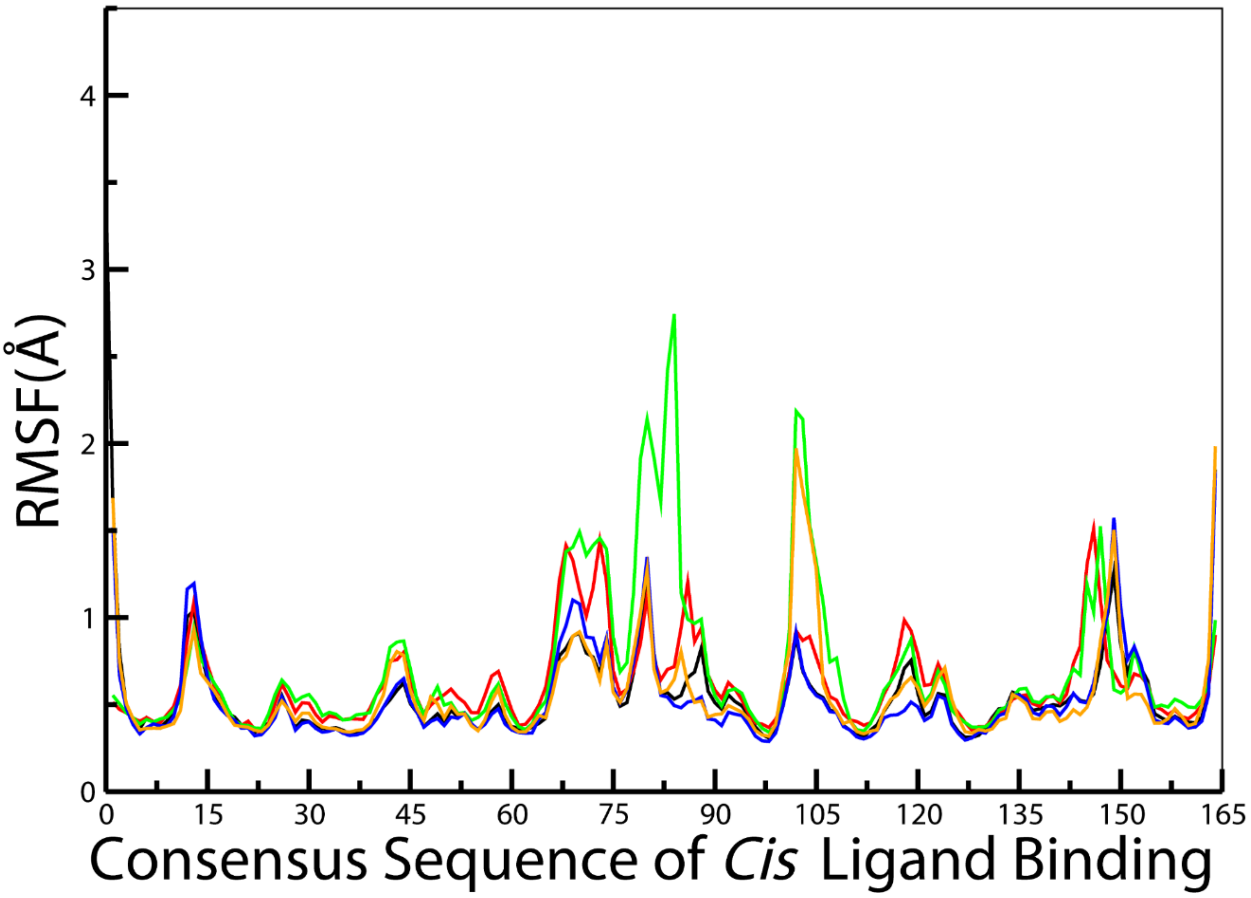


Figure 4-7 RMSF profiles of *Cis* Substrate Bound Cyclophilin Systems: CypA (black), CypB (red), CypC (green), CypD (blue), CypE (orange).
The RMSF for each system were initially ran on their entire protein sequence length. The output files were processed to align the residues of each system to that of CypA, in accordance to the individual sequence alignments.

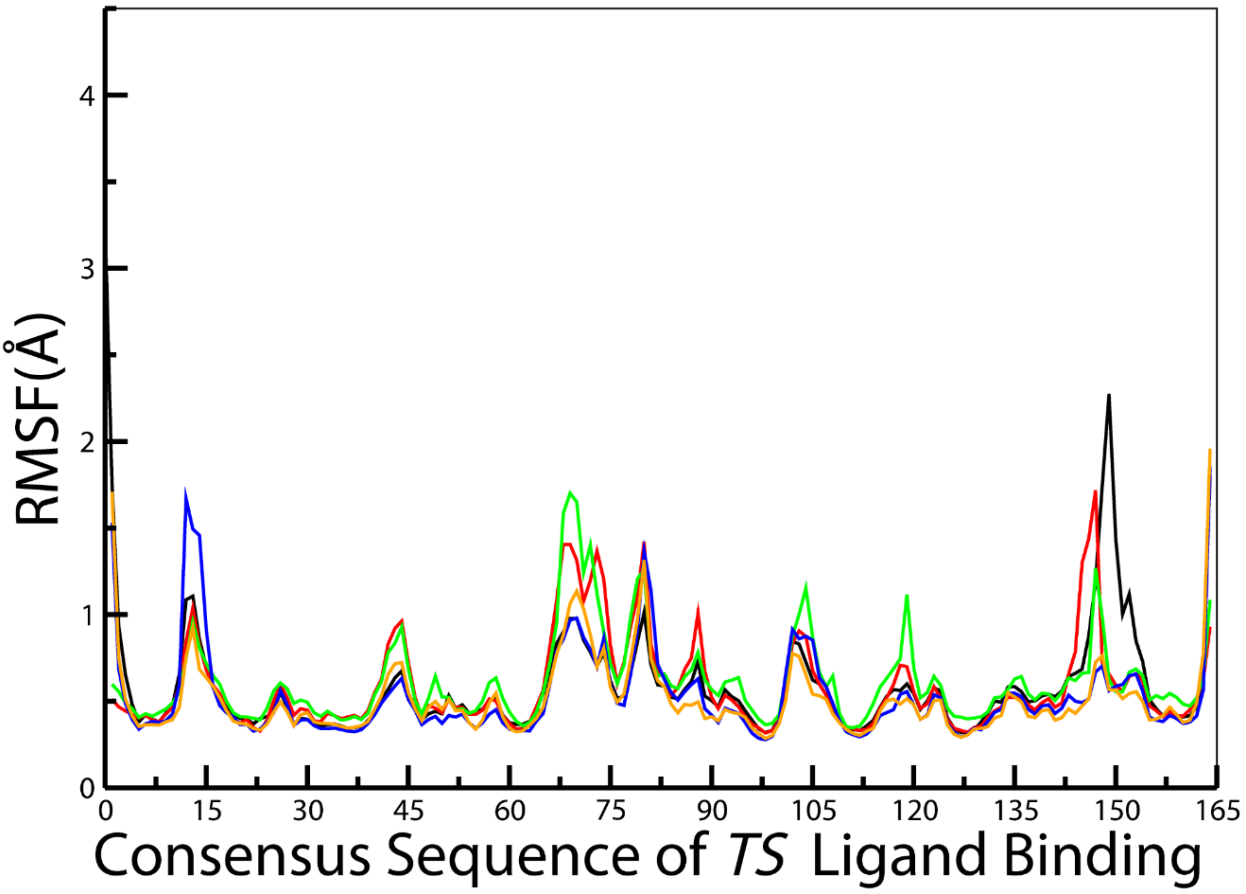


Figure 4-8 RMSF profiles of TS Substrate Bound Cyclophilin Systems: CypA (black), CypB (red), CypC (green), CypD (blue), CypE (orange).
The RMSF for each system were initially ran on their entire protein sequence length. The output files were processed to align the residues of each system to that of CypA, in accordance to the individual sequence alignments.

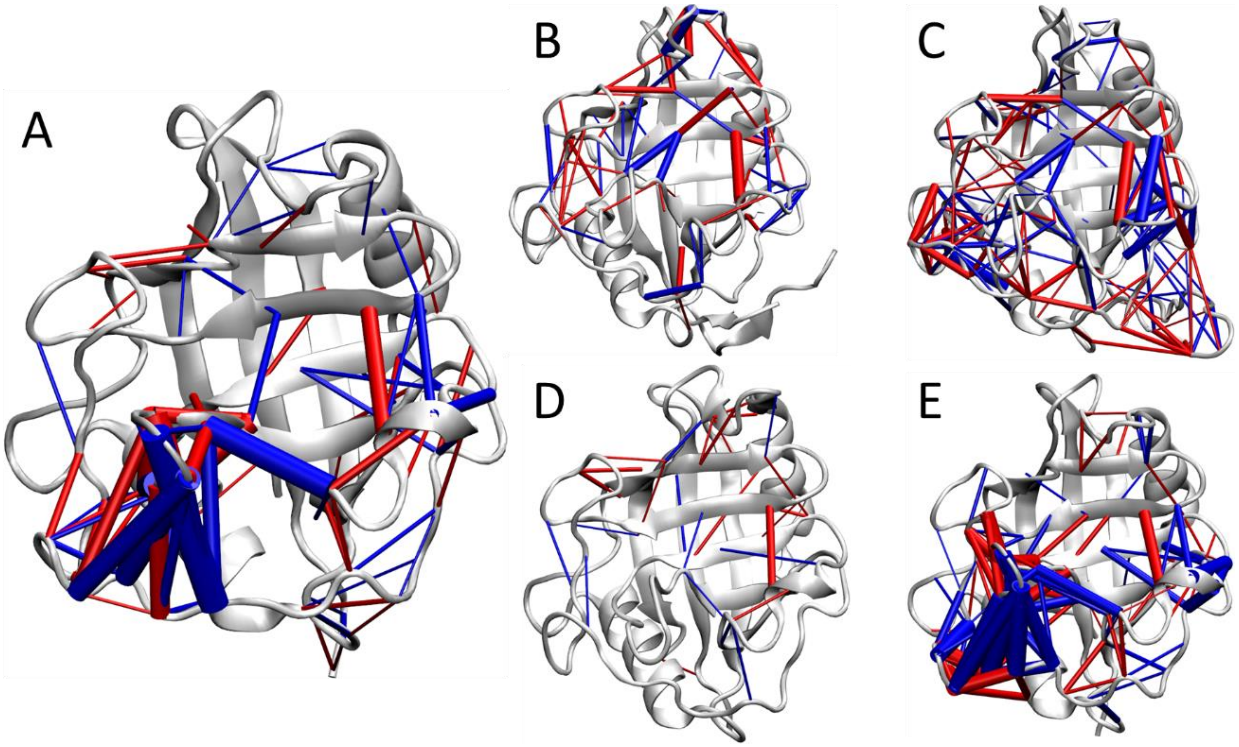


Figure 4-9 Contact Maps of Cyclophilin *Cis* Substrate Binding Dynamics.
The contact maps show a difference of more than 10% and less than 90% in the probability of contact formation or breaking when the free state binds to the substrate [$0.9 \leq abs(f_{cis}\text{-}f_{free}) \geq 0.1$] are shown as blue and red cylinders, respectively. The radius of the cylinders are proportional to the absolute values of $f_{cis}\text{-}f_{free}$.

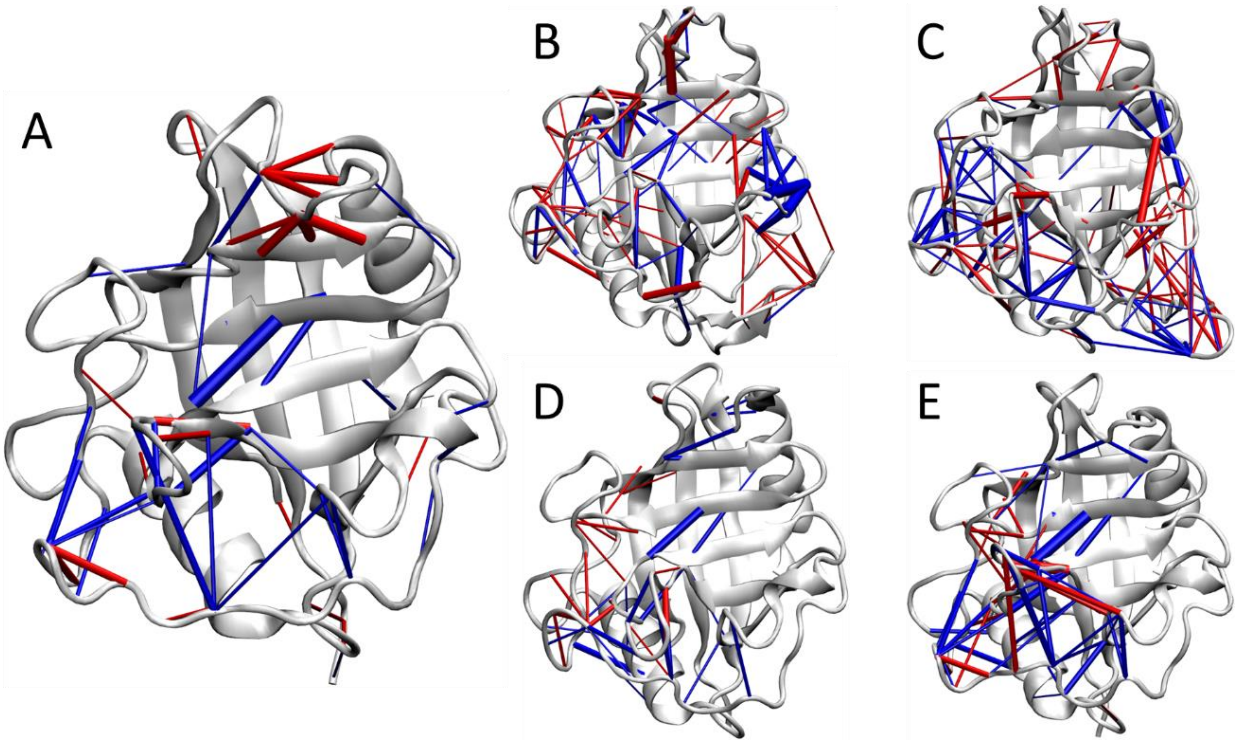


Figure 4-10 Contact Maps of the Transition of *Cis* Substrate Binding to *TS* Substrate Bound Dynamics.
The contact maps show a difference of more than 10% and less than 90% in the probability of contact formation or breaking when the *cis* substrate transitions to the *TS* substrate [$0.9 \leq abs(f_{cis}-f_{free}) \geq 0.1$] are shown as blue and red cylinders, respectively. The radius of the cylinders are proportional to the absolute values of $f_{cis}-f_{free}$.

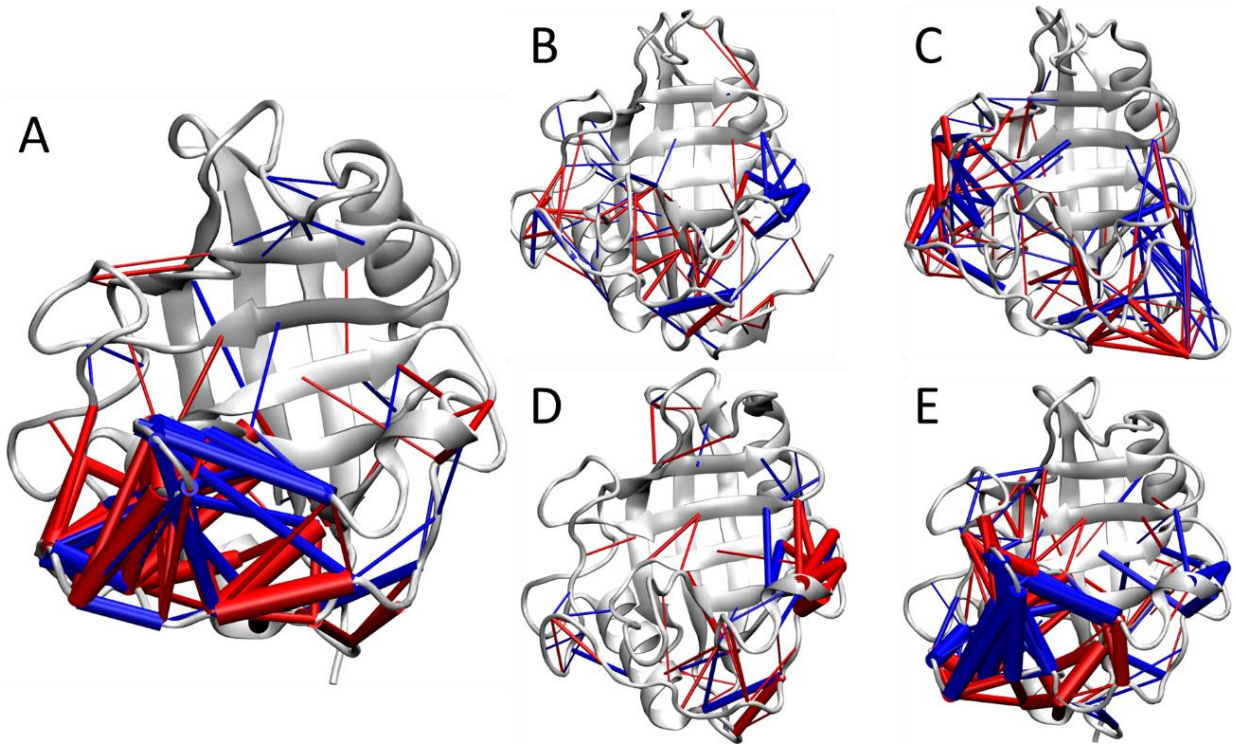


Figure 4-11 Contact Maps Valine to Leucine Mutants versus Wild Type Cyclophilin Dynamics.
The contact maps show a difference of more than 10% and less than 90% in the probability of contact formation or breaking when the wild-type cyclophilin becomes the mutant [$0.9 \leq abs(f_{cis}-f_{free}) \geq 0.1$] are shown as blue and red cylinders, respectively. The radius of the cylinders are proportional to the absolute values of $f_{cis}-f_{free}$.

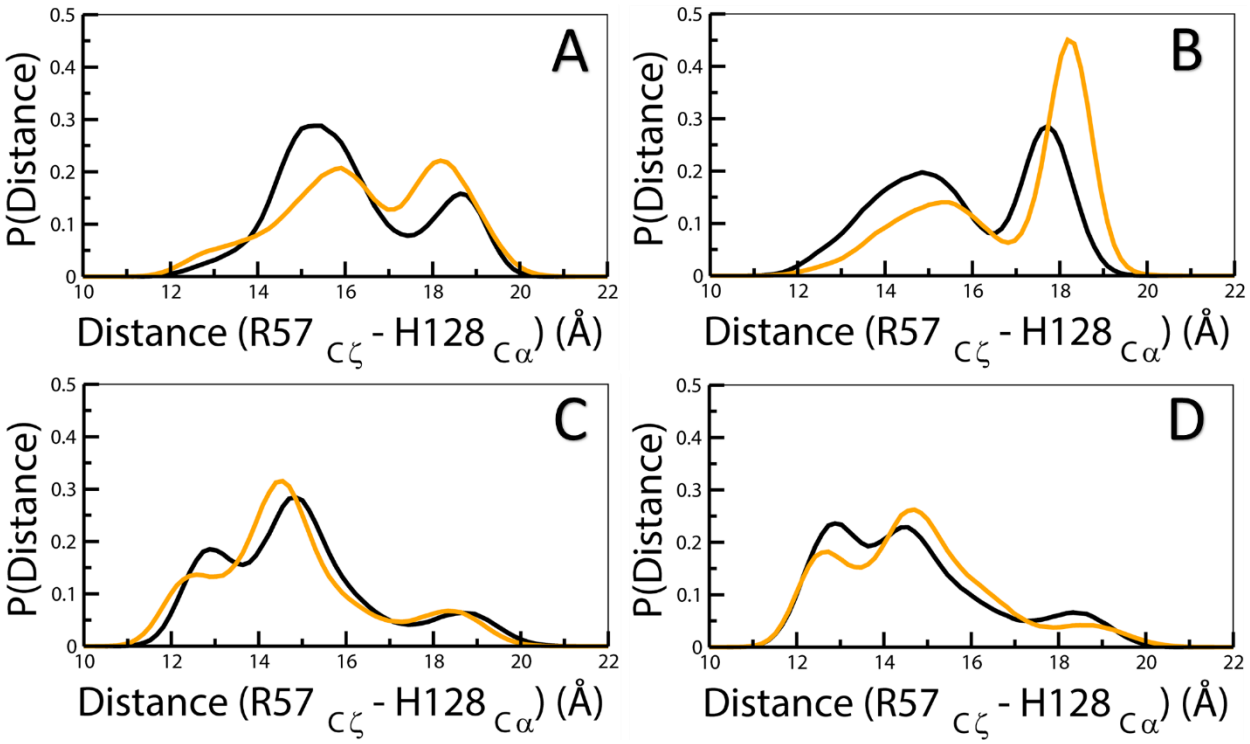


Figure 4-12 Distance Distribution between Catalytic Site Arginine and Histidine of Wild-Type and Valine to Leucine Mutant Cyclophilin Systems.
The distance was measured between the guanidinium carbon of the side chain of Arg-(CypB/CypC(57), CypD/CypE(54)) and backbone Cα atom of His-(CypB/CypC(128), CypD/CypE(125)) in the wild-type (black) and the mutant (orange). The distances for CypB (A) CypC (B), CypD (C), CypE (D) were represented as distributions.

4.2 Analysis of Cyclophilin Contact Dynamics

The contact dynamics of the cyclophilin isoforms were separated into two comparisons: (1) intra-systems: the comparison of the unbound, *cis* substrate bound, *TS* substrate bound, and mutant systems of the same homologue. (2) inter-system: the comparisons of the unbound isoforms to each other. These were performed to sample various combinations of comparative measures to test the similarity in the contact dynamics of each system. Contact PCA was the primary analysis used in the intra-system comparisons. Contact maps and PCA was used in the analysis of the inter-system comparisons.

4.2.1 Intra-System

The first two principal eigenvectors were projected for each isoform. The results of the contact PCA shows that in every isoform, there is a clear overlap in contact space of the *TS* substrate bound and *cis* substrate bound states. It was also evident that there was high overlapping of the contact space in all four states of CypB, CypC, and CypD. Of the homologues, only CypE exhibited similar separation of the unbound state from the bound and mutant states to that of CypA. This further suggests the dynamical conservation of CypE and CypA to be fairly high compared to the other homologues.

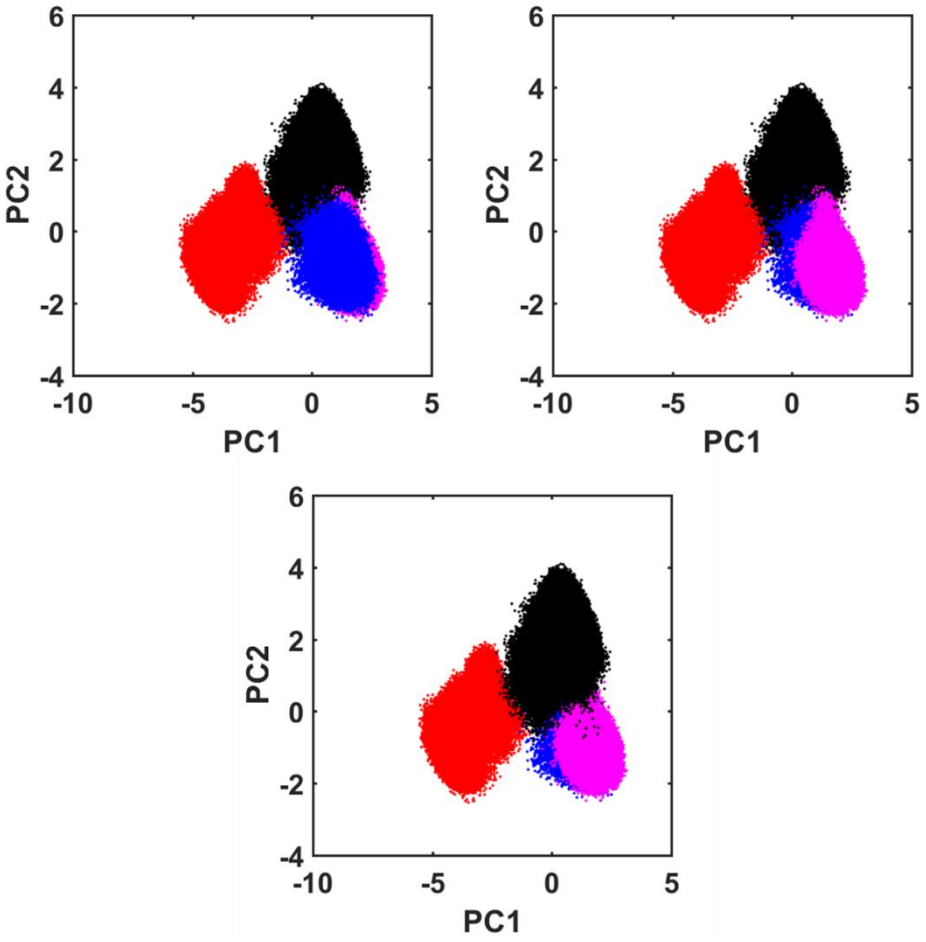


Figure 4-13 Contact PCA of Cyclophilin A Systems.
The first two principal components were projected for the contact space of the unbound (black), mutant (red), *cis* substrate bound (blue), and *TS* substrate bound (magenta) CypA.

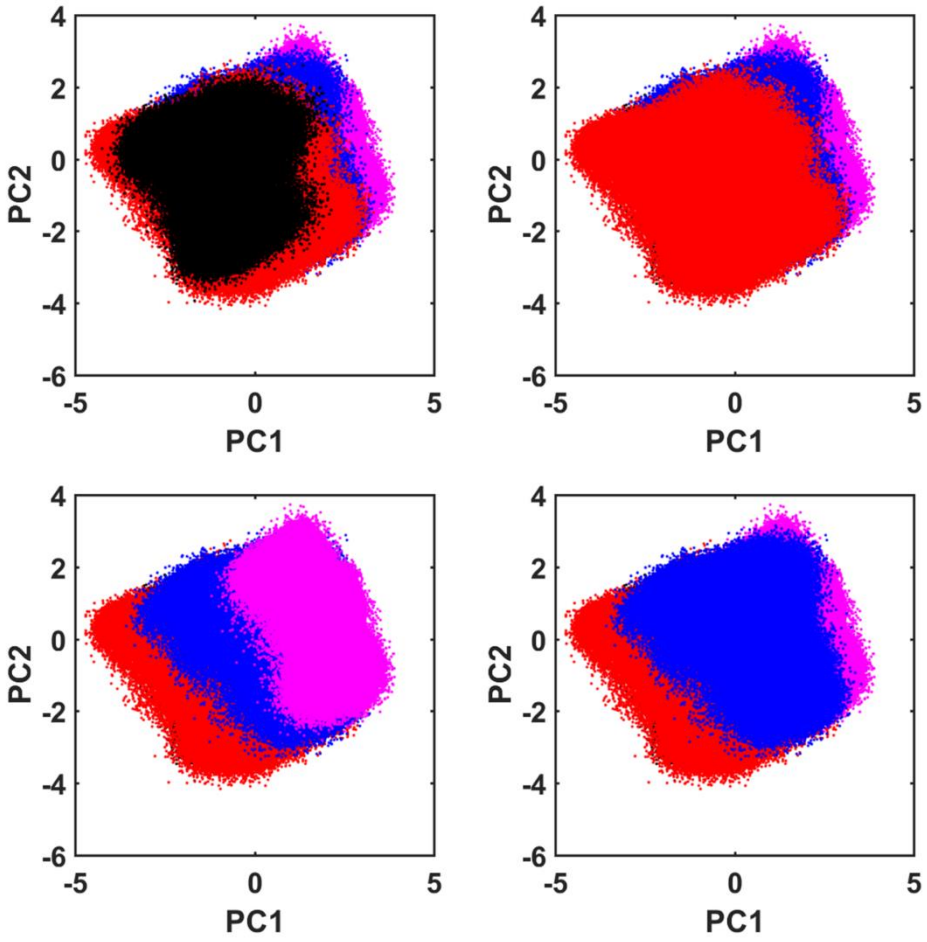


Figure 4-14 Contact PCA of Cyclophilin B Systems.
The first two principal components were projected for the contact space of the unbound (black), mutant (red), *cis* substrate bound (blue), and *TS* substrate bound (magenta) CypB.

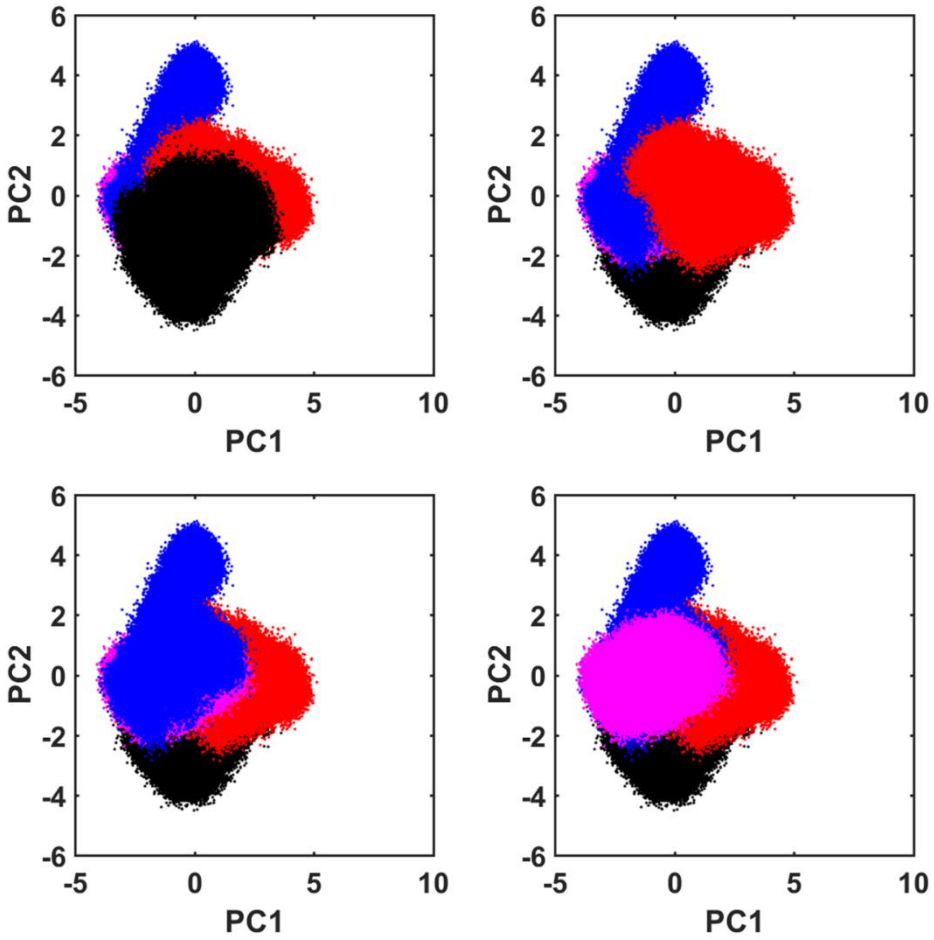


Figure 4-15 Contact PCA of Cyclophilin C Systems.
The first two principal components were projected for the contact space of the unbound (black), mutant (red), *cis* substrate bound (blue), and *TS* substrate bound (magenta) CypC.

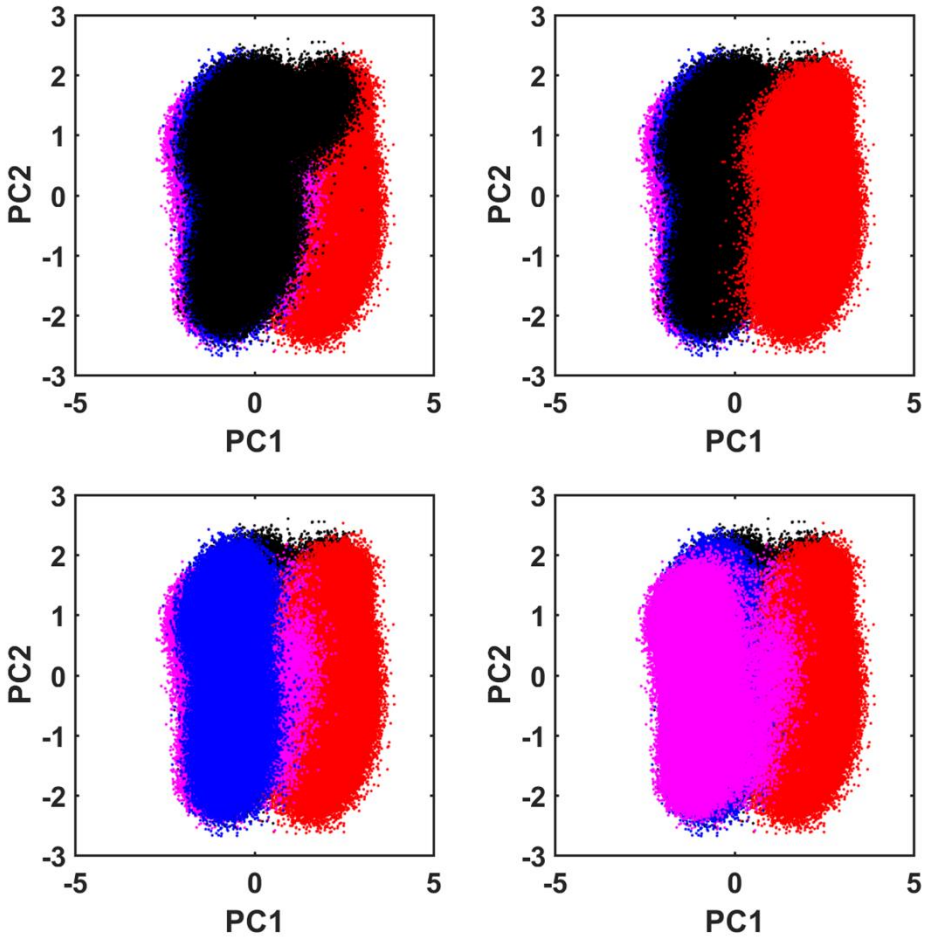


Figure 4-16 Contact PCA of Cyclophilin D Systems.
The first two principal components were projected for the contact space of the unbound (black), mutant (red), *cis* substrate bound (blue), and TS substrate bound (magenta) CypD.

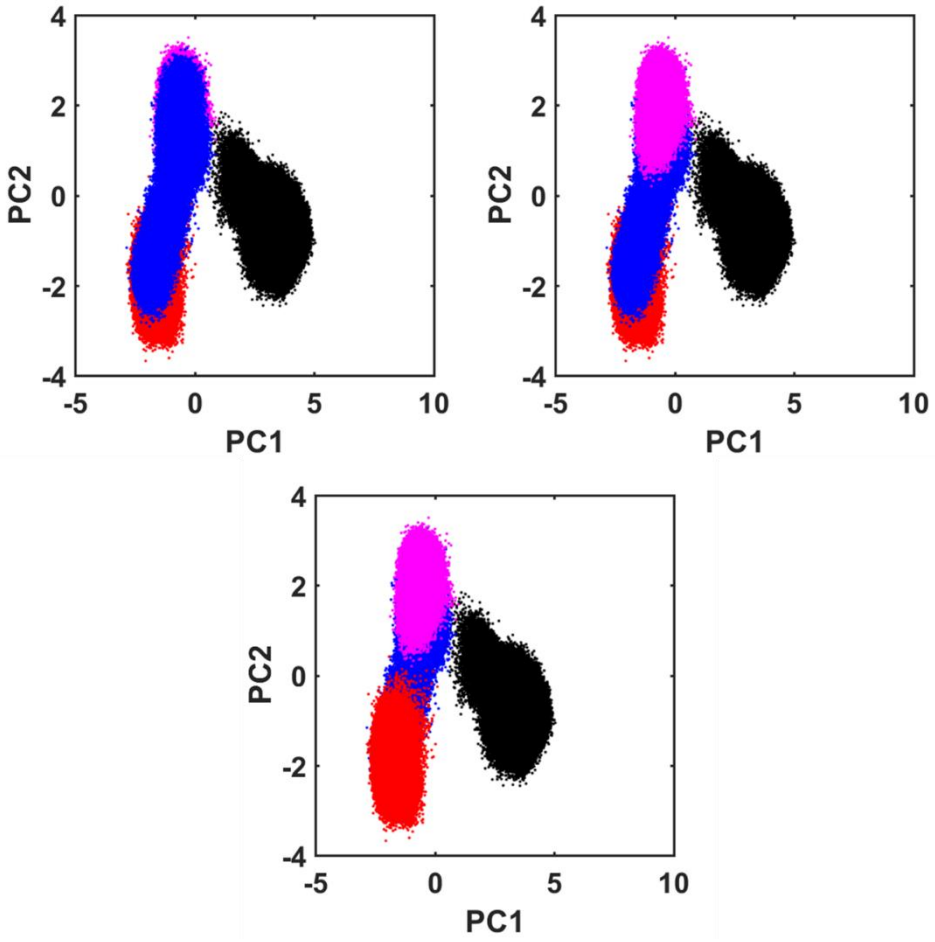


Figure 4-17 Contact PCA of Cyclophilin E Systems.
The first two principal components were projected for the contact space of the unbound (black), mutant (red), *cis* substrate bound (blue), and TS substrate bound (magenta) CypE.

4.2.2 Inter-System

The differences in contact dynamics between unbound CypA and its homologues were compared with three different methods. The first analysis directly compared the ranked contact indices of CypA and its homologues. The unbound state ranked contact index of each homologue was plotted against those of CypA. This was done to compare the sigmoidal region of each homologue to those of CypA. The initial thought was that the sigmoidal region of CypD would be slightly steeper than that of CypA, which would indicate its inflexibility in the unbound state. This was observed in Figure 4.18C. The curve of CypD ranked contacts were slightly steeper compared to that of CypA. The curves of CypB and CypC when compared to CypA exhibited similar trends. The curve of CypE was the one that was consistently close to that of CypA throughout the sigmoidal region. The contact PCA comparing the contact space occupied between the unbound CypA and its homologues were determined. As presented in Figure 4.19, there were clear separations between the CypA-B, CypA-C, and CypA-D comparisons, while there was some overlap in the contact space of CypA and CypE.

Contact maps were generated based on the contact differences between the unbound states of CypA and its homologues. This analysis was performed to determine the similarities in the discreet changes in contacts of the unbound states. The difference in contacts between CypA and CypB and CypC was expected to be significant and this was shown in Figure 4.20A-B. Based on the contact differences exhibited by CypD in Figure 3.10D, the difference between the unbound state of CypA and CypD were expected to significant as well. There was indeed a significant difference in the contacts of CypA and CypD. In contrast to CypD, the contact differences exhibited by CypE in Figure 4.10E were similar to that of CypA. If a difference were to be determined between the unbound states of CypA and CypE, one would expect the contacts

to cancel each other out. This was not the case as shown in Figure 4.20D. This actually revealed that even if a protein within the same family shares similar binding contact ensembles, the discreet and underlying contact dynamics of their unbound states can be vastly different from each other.

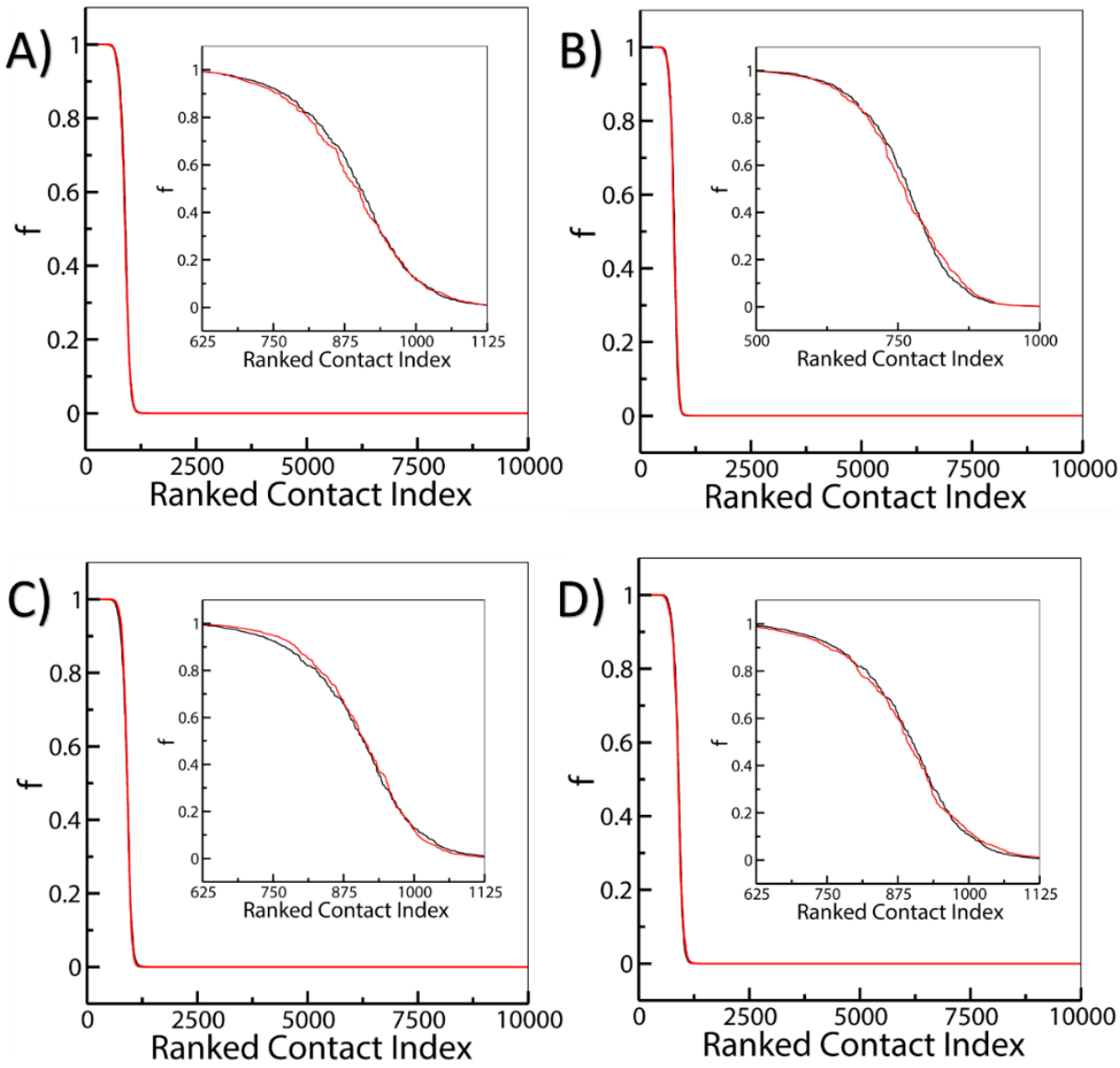


Figure 4-18 Ranked Contact Index Comparison of Unbound Cyclophilin Isoforms to Unbound Cyclophilin A.
The graphs are presented as followed: A-B (A), A-C (B), A-D (C), and A-E (D). The residue-residue contacts present in the cyclophilins were ranked based on their frequency of contact formation. The contact frequencies are plotted against their respective ranked contact index. The inner graph is a magnification of the contacts that are formed 10% to 90% of the simulated time.

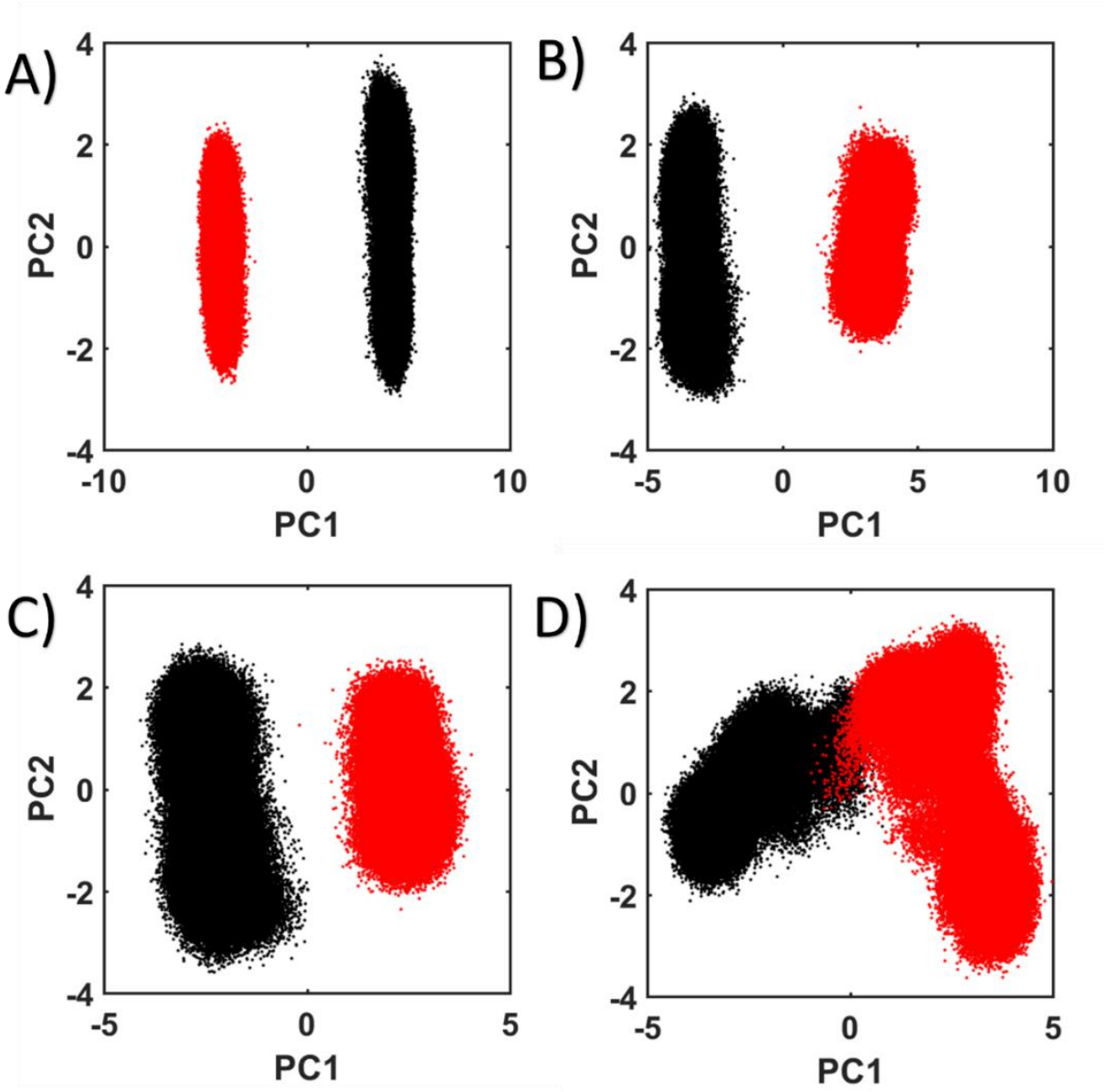


Figure 4-19 Contact PCA of Unbound Cyclophilin Homologues to Cyclophilin A.
The projection of the first two principal components of the contact space of CypB (A), CypC (B), CypD (C), and CypE (D) with CypA. The contact space of CypA are shown in black and the homologues are shown in red.

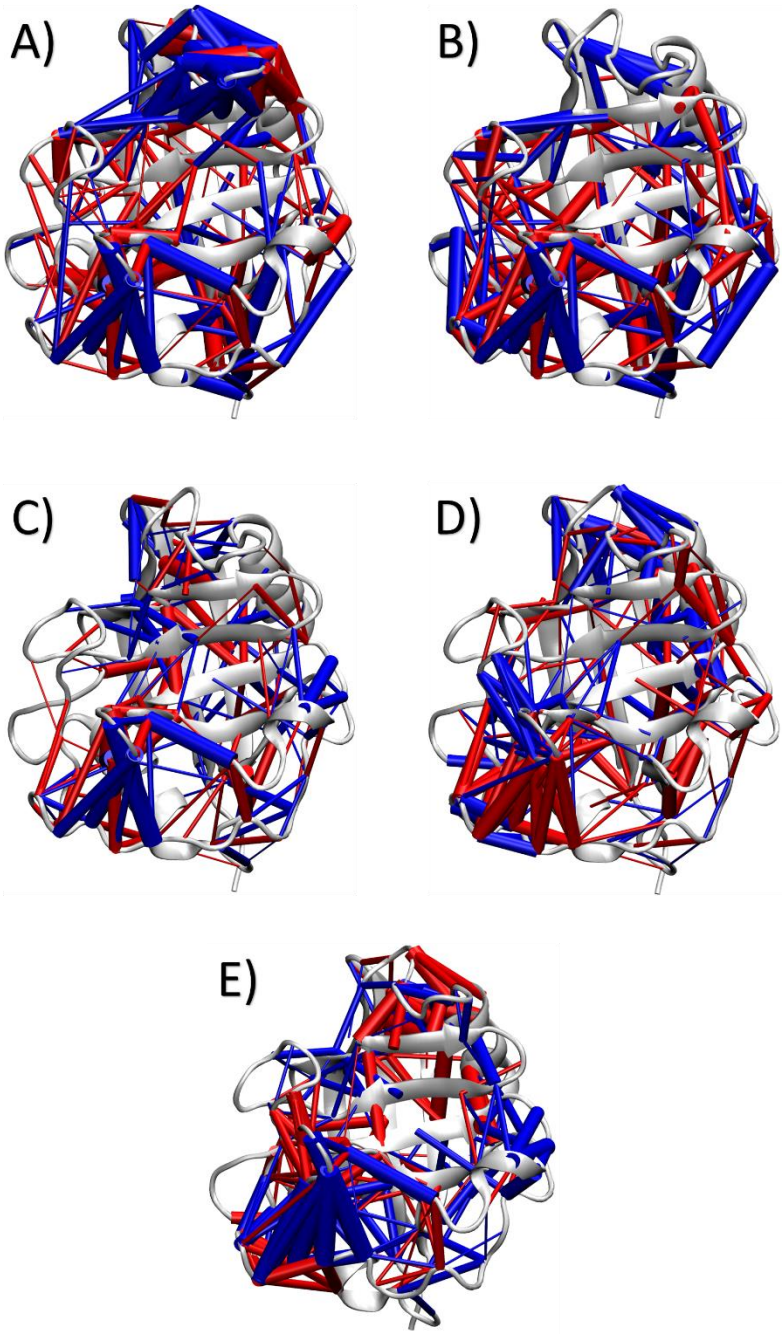


Figure 4-20 Contact Map of the Differences Between the Free State Contact Dynamics of Cyclophilin Isoforms.
The difference between CypB, CypC, CypD, and CypE to CypA are shown as A, B, C, and D, respectively. The difference between CypD and CypE is shown as E. The contact maps show a difference of more than 10% and less than 90% in the probability of contact formation or breaking when the wild-type cyclophilin becomes the mutant [$0.9 \leq abs(f_{cis}\text{-}f_{free}) \geq 0.1$] are shown as blue and red cylinders, respectively. The radius of the cylinders are proportional to the absolute values of $f_{cis}\text{-}f_{free}$.

4.3 Cyclophilin Homologue Dynamical Conservation to Cyclophilin A

Comparative analysis of the cyclophilin isoforms were performed in order to determine their dynamical conservation and to what extent are they conserved. The first comparative measure was the backbone Cartesian PCA analysis between CypA and its homologues. The projection of the conformational space of the cyclophilin isoforms presented in Figure 4.21 were similar to that of the contact space comparison in Figure 4.19. There are clear separation in the CypA-B, CypA-C, and CypA-D comparisons while there are significant overlap of the conformational space of CypA and CypE.

The contact differences of CypA and its homologues were plotted against each other to generate plots of pairwise contact coordinates. Due to previous results of CypE and its similarity to CypA, the CypAE difference plot was used to define the dynamical conservation of the cyclophilin isoforms. As shown in Figure 4.25, the majority of pairwise contacts were located in the top right and bottom left quadrants of the plot. This shows that contacts in two different isoforms that change in the same direction, either a positive-positive or a negative-negative change, can be considered positively correlated. The unbound → *cis* differences shown in Figures 4.22-4.24, indicates that there's a level of dynamical conservation in the CypAB and CypAC differences while the CypAD indicates that there is little to no dynamical conservation. This was expected due to the unbound → *cis* contact map of CypD, presented in Figure 4.9. As stated back in section 4.1, conservation of contact dynamics in cyclophilins can be observed in their catalytic dynamics. The *cis* → *TS* differences shown in Figures 4.22-4.25, indicates a shift in pairwise contacts into the correlated regions.

The contact differences of CypD was also plotted against those of CypE, as presented in Figure 4.26. The plots were very similar to those of CypAD shown in Figure 4.24. This lead to

the further investigation of the contacts of CypA, CypD, and CypE. Pairwise contacts that are present in CypA, CypD, and CypE from the dynamically conserved regions of each plot, CypAD, CypAE, and CypED, were extracted and shown in Tables 4.1-4.2. Among all the contacts, the contact between residues 101-103 in CypA and residues 100-102 in CypD and CypE were present in both the unbound → *cis* and *cis* → *TS* differences. In addition, this contact was located in the negative-negative dynamically conserved region, which means that it is a contact that needs to be broken during the initial binding of a substrate and the catalysis of the substrate. Another interesting dynamic that was revealed in the dynamically conserved contacts was residue-residue contact pairs that were either broken during substrate binding but formed during catalysis or. As shown in Tables 4.1-4.2, the contact between residues 101-108, 88-126, 62-112, and 61-64 in CypA and residues 100-107, 87-125, 61-111, and 61-63 in CypD and CypE were broken during substrate binding but formed during catalysis. As shown in Tables 4.1-4.2, the contact between residues 77-81 and 65-74 in CypA and residues 76-80 and 64-73 in CypD and CypE were formed during substrate binding but broken during catalysis. Aside from these contact pairs, the rest of the contact pairs between Table 4.1 and Table 4.2 are unique. This suggests that during catalysis, CypA, CypD, and CypE adopts a new set of contacts that is different to the contacts that occur during substrate binding.

The dynamical conservation indices of every possible combinations of cyclophilin A, B, C, D, and E were calculated based on the average of the translated data points defined in section 3.5.5. The indices are presented in Tables 4.3-4.4. The dynamical conservation indices of the unbound → *cis* shown in Table 4.3 indicates that the dynamical conservation of CypE to CypA was the highest. There is an overall increase in dynamical conservation indices when the unbound → *cis* values move towards the *cis* → *TS* values. One abnormality was the *cis* → *TS*

value between CypB and CypC, which reported a substantial decrease in dynamical conservation.

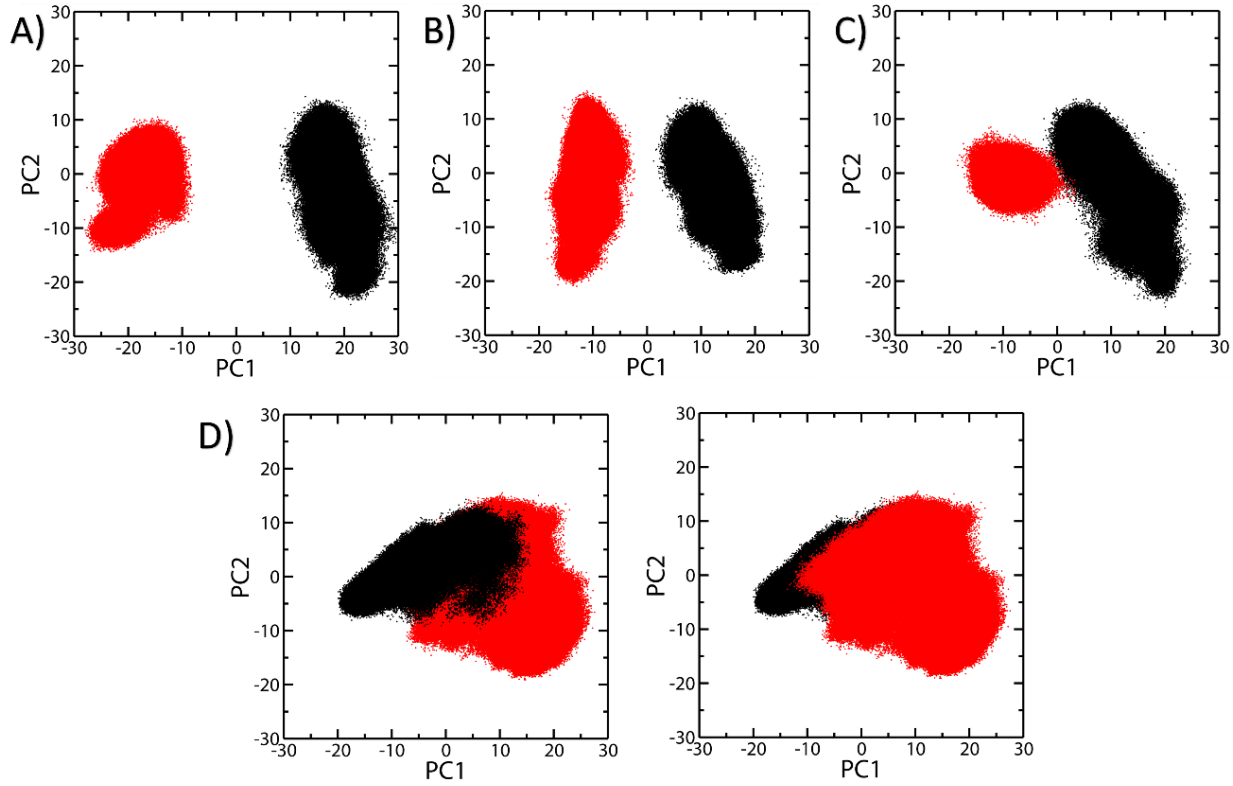


Figure 4-21 Protein Backbone Cartesian PCA of Unbound Cyclophilin Isoforms CypB (A), CypC (B), CypD (C), CypE (D) compared to those of CypA.
The first two principal components were projected for the internal backbone motion of CypA (black) and homologues (red).

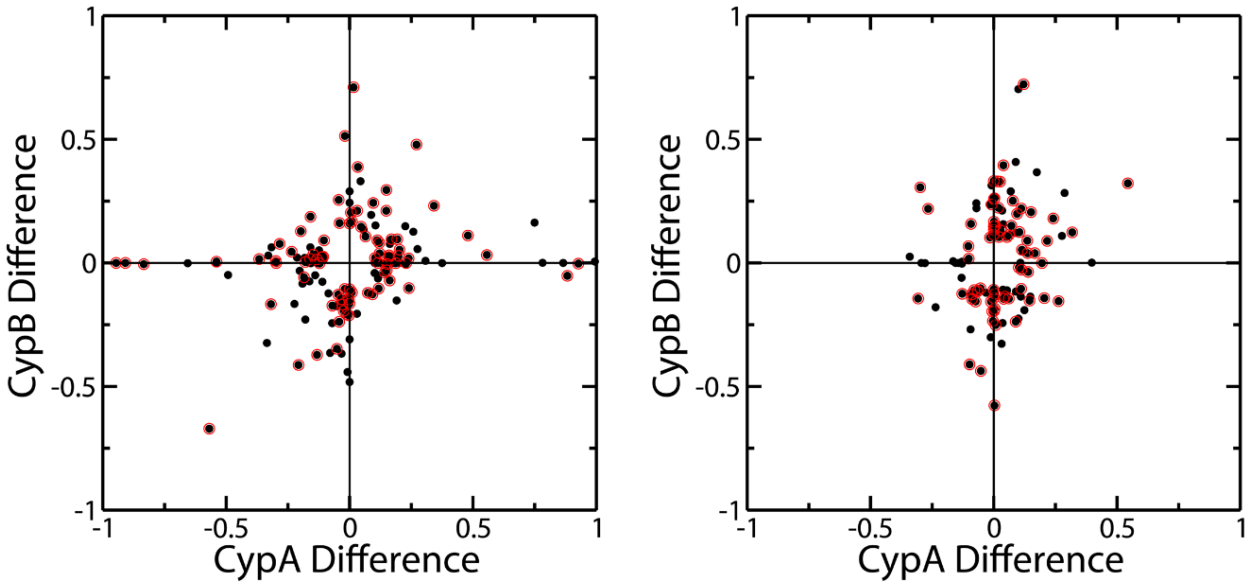


Figure 4-22 Plotted Unbound → *Cis* (left) and *Cis* → *TS* (right) Contact Differences of CypB against CypA.

The contact statistics of CypB was aligned to that of CypA and the difference of the *cis* substrate bound state and free states were determined. Further refinement of the data set was performed by eliminating data values that fit the 0.1 cut-off. Those difference values were plotted as black dots. The contact pairs that remained after the cutoff were screened for conservation and separated based on matching residue to residue contacts. The conserved residue pairs are indicated with a red ring.

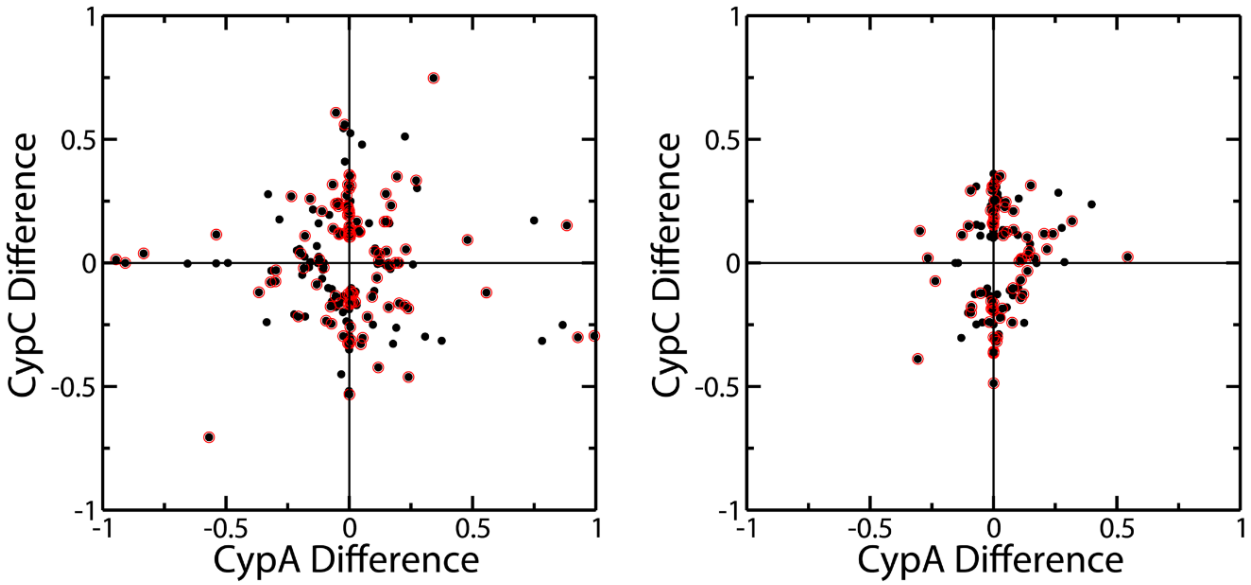


Figure 4-23 Plotted Unbound → *Cis* (left) and *Cis* → *TS* (right) Contact Differences of CypC against CypA.

The contact statistics of CypC was aligned to that of CypA and the difference of the *cis* substrate bound state and free states were determined. Further refinement of the data set was performed by eliminating data values that fit the 0.1 cut-off. Those difference values were plotted as black dots. The contact pairs that remained after the cutoff were screened for conservation and separated based on matching residue to residue contacts. The conserved residue pairs are indicated with a red ring.

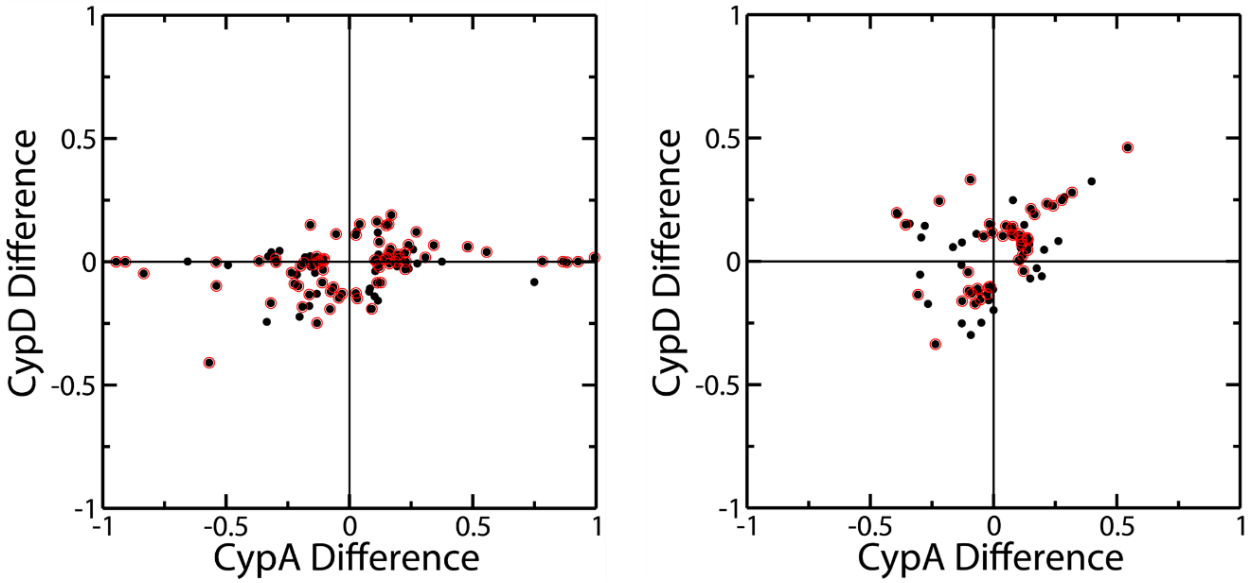


Figure 4-24 Plotted Unbound → *Cis* (left) and *Cis* → *TS* (right) Contact Differences of CypD against CypA.

The contact statistics of CypD was aligned to that of CypA and the difference of the *cis* substrate bound state and free states were determined. Further refinement of the data set was performed by eliminating data values that fit the 0.1 cut-off. Those difference values were plotted as black dots. The contact pairs that remained after the cutoff were screened for conservation and separated based on matching residue to residue contacts. The conserved residue pairs are indicated with a red ring.

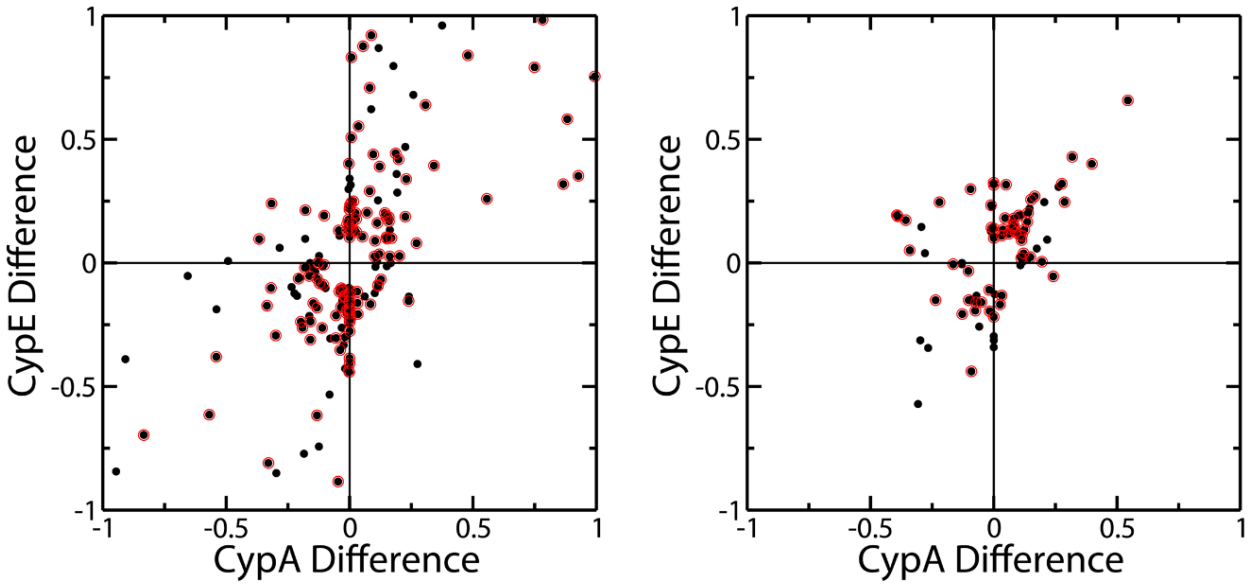


Figure 4-25 Plotted Unbound → *Cis* (left) and *Cis* → *TS* (right) Contact Differences of CypE against CypA.

The contact statistics of CypE was aligned to that of CypA and the difference of the *cis* substrate bound state and free states were determined. Further refinement of the data set was performed by eliminating data values that fit the 0.1 cut-off. Those difference values were plotted as black dots. The contact pairs that remained after the cutoff were screened for conservation and separated based on matching residue to residue contacts. The conserved residue pairs are indicated with a red ring.

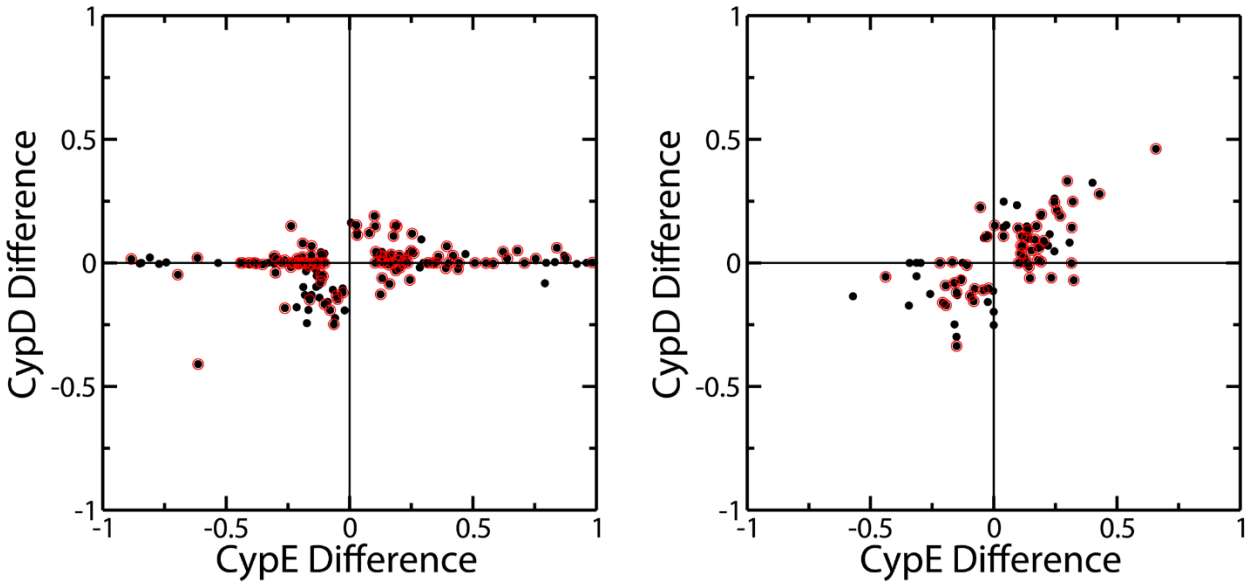


Figure 4-26 Plotted Unbound → *Cis* (left) and *Cis* → *TS* (right) Contact Differences of CypD against CypE.

The contact statistics of CypD was aligned to that of CypE and the difference of the *cis* substrate bound state and free states were determined. Further refinement of the data set was performed by eliminating data values that fit the 0.1 cut-off. Those difference values were plotted as black dots. The contact pairs that remained after the cutoff were screened for conservation and separated based on matching residue to residue contacts. The conserved residue pairs are indicated with a red ring.

Table 4-1 Dynamically Conserved CypA, CypD, and CypE Residue-Residue Contacts of Unbound → *Cis.*

The residue-residue contacts that were present in the isoforms were extracted from the dynamical conservation plots. The contact pairs were color coated based on interaction type. Conserved residue pair interactions with respect to CypA are shown in black. Polar negative (P^-) shift to a polar neutral (P°) and vice versa are shown in red. Polar positive (P^+) shift to a P^- and vice versa are shown in green. An aliphatic shift to a polar and vice versa are shown in blue. The strengths refer to the difference in the unbound → *cis* contacts.

	A	Strength	D	Strength	E	Strength
Positive-Positive Differences	LYS28 - ASN87	0.1905	VAL27 - ASN86	0.02560	VAL27 - ASN86	0.3596
	ARG55 - MET61	0.1486	ARG54 - MET60	0.1478	ARG54 - MET60	0.1911
	ILE57 - ARG148	0.1144	ILE56 - LYS147	0.1187	ILE56 - LYS147	0.2533
	PHE60 - TRP121	0.3416	PHE59 - TRP120	0.06700	PHE59 - TRP120	0.3939
	PHE60 - LEU122	0.1426	PHE59 - LEU121	0.00957	PHE59 - LEU121	0.2017
	GLN63 - ALA101	0.2712	GLN62 - ALA100	0.1211	GLN62 - ALA100	0.07989
	GLY65 - GLY74	0.02596	GLY64 - GLY73	0.1086	GLY64 - GLY73	0.1781
	THR68 - LYS76	0.1492	THR67 - LYS75	0.1476	THR67 - LYS75	0.1037
	SER77 - GLU81	0.1178	SER76 - SER80	0.02983	SER76 - LYS80	0.8696
	PHE83 - ASN106	0.7824	PHE82 - ASN105	0.00161	PHE82 - ASP84	0.9841
	GLU84 - ASN106	0.3746	PRO83 - ASN105	0.00017	ASP83 - ASN105	0.9608
	GLU84 - ASN108	0.1775	PRO83 - ASN107	0.00019	ASP83 - ASN107	0.7969
	ASP85 - ASN102	0.5565	ASP84 - ASN101	0.03980	ASP84 - ASN101	0.2593
	ASP85 - PRO105	0.9937	ASP84 - PRO104	0.01702	ASP84 - PRO104	0.7543
	ASP85 - ASN106	0.3078	ASP84 - ASN105	0.01791	ASP84 - ASN105	0.6388
	PHE88 - LYS91	0.1594	PHE87 - LYS90	0.03463	PHE87 - LYS90	0.1688
	HID92 - PHE113	0.1984	HID91 - PHE112	0.02908	HID91 - PHE112	0.4192
	HID92 - GLU120	0.2582	HID91 - ASP119	0.04958	HID91 - ASP119	0.6803
	ASN102 - HIP126	0.2295	ASN101 - HIP125	0.00007	ASN101 - HIP125	0.3394
	ASN106 - ASN108	0.8655	ASN105 - ASN107	0.00082	ASN105 - ASN107	0.3185
	THR107 - GLY109	0.9271	THR106 - GLY108	0.00092	THR106 - GLY108	0.3519
	PHE113 - THR119	0.1590	PHE112 - THR118	0.1531	PHE112 - THR118	0.1848
	CYS115 - THR119	0.2257	CYS114 - THR118	0.03603	THR114 - THR118	0.4696
	LYS118 - TRP121	0.4795	LYS117 - TRP120	0.06092	LYS117 - TRP120	0.8401
	LYS125 - VAL127	0.1699	LYS124 - VAL126	0.1897	LYS124 - VAL126	0.1005
Negative-Negative Differences	GLY50 - HIP70	-0.1318	GLY49 - HIP69	-0.2485	GLY49 - HIP69	-0.06386
	CYS52 - HIP70	-0.2024	THR51 - HIP69	-0.2232	SER51 - HIP69	-0.05887
	HIE54 - ASN71	-0.1632	HIE53 - ASN70	-0.1332	HIE53 - ASN70	-0.05266
	ARG55 - ASN149	-0.1629	ARG54 - SER148	-0.1794	ARG54 - ASP148	-0.2146
	ARG55 - LYS151	-0.1321	ARG54 - ARG150	-0.1300	ARG54 - LYS150	-0.1809
	ILE57 - THR116	-0.03008	ILE56 - THR115	-0.1293	ILE56 - CYS115	-0.1548
	MET61 - LEU122	-0.5688	MET60 - LEU121	-0.4095	MET60 - LEU121	-0.6141
	CYS62 - GLY64	-0.3347	CYS61 - ALA63	-0.2434	CYS61 - GLY63	-0.1728
	CYS62 - PHE112	-0.1912	CYS61 - PHE111	-0.1828	CYS61 - PHE111	-0.2622
	THR73 - ALA103	-0.1848	THR72 - ALA102	-0.004113	THR72 - SER102	-0.7718
	PHE88 - HIP126	-0.1977	PHE87 - HIP125	-0.01676	PHE87 - HIP125	-0.2381
	HID92 - ALA117	-0.2130	HID91 - ILE116	-0.05194	HID91 - ASP116	-0.1330
	ALA101 - ALA103	-0.5397	ALA100 - ALA102	-0.09771	ALA100 - SER102	-0.1878
	ALA101 - ASN108	-0.8341	ALA100 - ASN107	-0.04765	ALA100 - ASN107	-0.6960
	ASN102 - ASN106	-0.5404	ASN101 - ASN105	-0.001776	ASN101 - ASN105	-0.3797
	ALA103 - ASN108	-0.9462	ALA102 - ASN107	-0.000002	SER102 - ASN107	-0.8434
	ALA103 - GLN111	-0.2971	ALA102 - GLN110	-0.001911	SER102 - GLN110	-0.8502
	CYS115 - LEU122	-0.2237	CYS114 - LEU121	-0.08884	THR114 - LEU121	-0.1232
	THR119 - ASP123	-0.1474	THR118 - ASP122	-0.01168	THR118 - ASP122	-0.1631
	TRP121 - LYS125	-0.3187	TRP120 - LYS124	-0.1674	TRP120 - LYS124	-0.1014

Table 4-2 Dynamically Conserved CypA, CypD, and CypE Residue-Residue Contacts of *Cis* → *TS* .

The residue-residue contacts that were present in the isoforms were extracted from the dynamical conservation plots. The contact pairs were color coated based on interaction type. Conserved residue pair interactions with respect to CypA are shown in black. P^- shift to a P° and vice versa are shown in red. P^+ shift to a P^- and vice versa are shown in green. An aliphatic shift to a polar and vice versa are shown in blue. The strengths refer to the difference in the *Cis* → *TS* contacts.

	A	Strength	D	Strength	E	Strength
Positive-Positive Differences	HIE54 - ILE56	0.2874	HIE53 - VAL55	0.2606	HIE53 - ILE55	0.2463
	HIE54 - GLN111	0.1144	HIE53 - GLN110	0.07026	HIE53 - GLN110	0.1109
	ARG55 - PHE60	0.1042	ARG54 - PHE59	0.1100	ARG54 - PHE59	0.1137
	CYS62 - GLY64	0.3979	CYS61 - ALA63	0.3244	CYS61 - GLY63	0.4003
	CYS62 - PHE112	0.2766	CYS61 - PHE111	0.2482	CYS61 - PHE111	0.3207
	CYS62 - GLU143	0.03767	CYS61 - GLU142	0.1029	CYS61 - GLU142	0.1351
	GLN63 - GLN111	0.54407	GLN62 - GLN110	0.4617	GLN62 - GLN110	0.6576
	GLY75 - GLU81	0.1449	GLY74 - SER80	0.07018	GLY74 - LYS80	0.2203
	LYS76 - GLY80	0.1018	LYS75 - GLY79	0.005038	LYS75 - GLY79	0.1919
	LYS76 - GLU81	0.2628	LYS75 - SER80	0.0827	LYS75 - LYS80	0.3078
	SER77 - SER110	0.07140	SER76 - SER109	0.1313	SER76 - SER109	0.1377
	GLY80 - LYS82	0.1242	GLY79 - ARG81	0.1481	GLY79 - LYS81	0.1347
	GLU81 - ASN108	0.1112	SER80 - ASN107	0.06777	LYS80 - ASN107	0.1938
	GLU81 - GLY109	0.2052	SER80 - GLY108	0.04690	LYS80 - GLY108	0.2460
	ASP85 - ASN102	0.1511	ASP84 - ASN101	0.2130	ASP84 - ASN101	0.2559
	ASP85 - LYS125	0.1361	ASP84 - LYS124	0.09476	ASP84 - LYS124	0.1664
	GLU86 - VAL127	0.05005	GLU85 - VAL126	0.1431	GLU85 - VAL126	0.3166
	PHE88 - LYS125	0.1666	PHE87 - LYS124	0.1907	PHE87 - LYS124	0.2695
	PHE88 - HIP126	0.1398	PHE87 - HIP125	0.08959	PHE87 - HIP125	0.2028
	ALA101 - ASN108	0.3184	ALA100 - ASN107	0.2791	ALA100 - ASN107	0.4290
Negative-Negative Differences	ASP13 - LYS155	-0.1287	ASN12 - LYS154	-0.2513	GLY12 - LYS154	-0.000105
	LEU39 - GLY109	-0.1276	LEU38 - GLY108	-0.1610	LEU38 - GLY108	-0.2065
	TYR48 - PHE67	-0.08909	TYR47 - PHE66	-0.1302	PHE47 - PHE66	-0.1479
	CYS52 - ASP66	-0.05893	THR51 - ASP65	-0.1258	SER51 - ASP65	-0.2573
	GLY65 - GLY74	-0.07492	GLY64 - GLY73	-0.1715	GLY64 - GLY73	-0.1934
	GLY74 - SER110	-0.1012	GLY73 - SER109	-0.1193	GLY73 - SER109	-0.1509
	SER77 - GLU81	-0.09225	SER76 - SER80	-0.2990	SER76 - LYS80	-0.1511
	TYR79 - GLU81	-0.2660	TYR78 - SER80	-0.1724	TYR78 - LYS80	0.1347
	GLU81 - PHE83	-0.2979	SER80 - PHE82	-0.05383	LYS80 - PHE82	-0.3133
	LYS82 - GLY109	-0.04956	ARG81 - GLY108	-0.2485	LYS81 - GLY108	-0.1589
	ALA101 - ALA103	-0.3068	ALA100 - ALA102	-0.1349	ALA100 - SER102	-0.5711
	ASN102 - PRO105	-0.2361	ASN101 - PRO104	-0.3361	ASN101 - PRO104	-0.1508

Table 4-3 Dynamical Conservation Index of *Cis* Substrate Binding Contacts of Cyclophilin Isoforms.

	A	B	C	D	E
A		0.2230	-0.07568	0.2788	0.5667
B			0.5050	0.2766	-0.06915
C				0.1149	-0.09589
D					0.40625
E					

Table 4-4 Dynamical Conservation Index of the Transition of the *Cis* Substrate Bound State to the *TS* Substrate Bound State Contacts of Cyclophilin Isoforms.

	A	B	C	D	E
A		0.2137	0.09091	0.4722	0.4337
B			0.09596	0.2881	0.2595
C				-0.05600	0.1037
D					0.6000
E					

5 CONCLUSIONS

Previous MD simulations done on CypA revealed an interesting set of dynamics when it is bound to a substrate. When binding a *cis* substrate, CypA exhibits significant contact ensemble at a site ~15Å away from the catalytic site. This special contact ensemble was predicted to show up in all cyclophilin isoforms. Through MD simulations, we were able to define the dynamics of the CypA homologues CypB, CypC, CypD, and CypE. To our surprise, the contact ensemble was only exhibited in CypE. In addition to CypE, CypD also exhibited an interesting contact ensemble or lack thereof. This lead to further MD simulations, which defined the dynamics of the *TS* substrate bound cyclophilin isoforms. The contact dynamics of the *TS* substrate bound cyclophilin isoforms revealed that the contact ensembles for all the simulated isoforms became more similar to each other. This suggests that it is the not the initial binding dynamics that are conserved, but the dynamics of the catalysis that are conserved.

Through the information revealed by the *TS* substrate bound cyclophilin isoforms, we were able to perform further comparative studies on the isoforms. This time, instead of only intra-system to intra-system comparisons, we performed direct inter-system comparisons. We were able to determine that CypA does not share contact and conformational space with CypB, CypC, and CypD, but does share contact and conformational space with CypE. We defined the dynamical conservation of the cyclophilin isoforms and identified various residue-residue contacts that are conserved between CypA, CypD, and CypE. These residue-residue contacts could be used in future MD studies involving mutations of specific residues in one isoform in order alter its dynamics to mimic those of another isoform.

REFERENCES

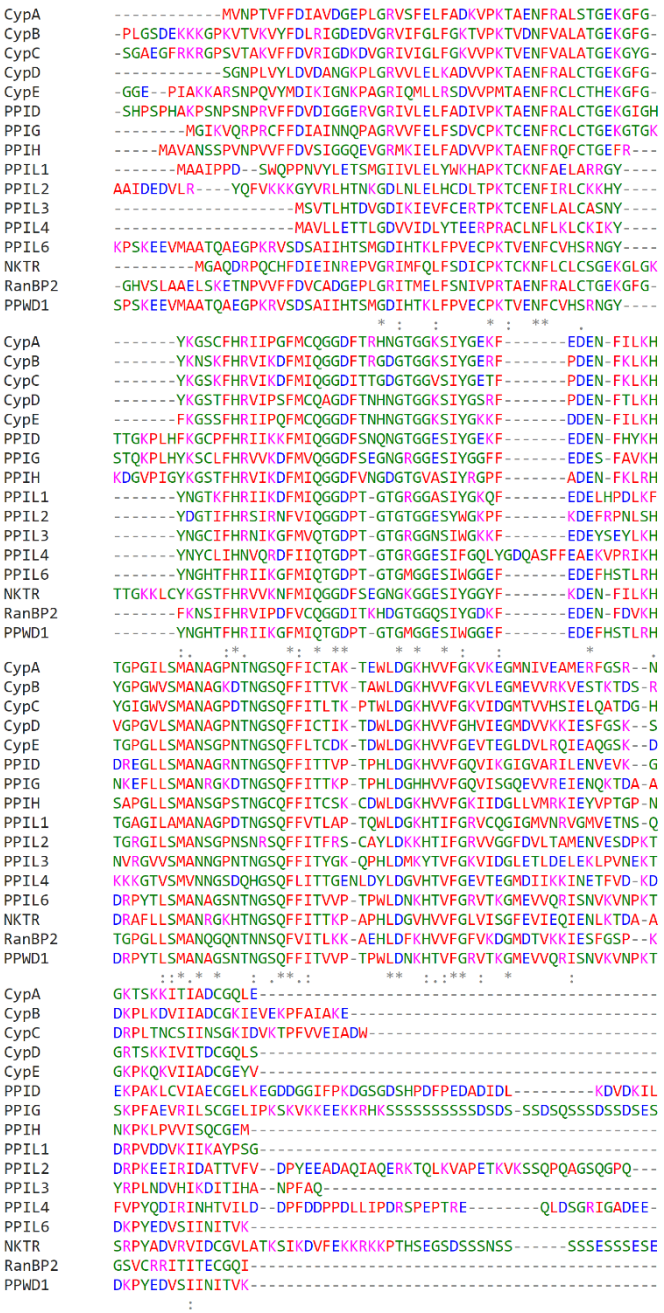
1. Chou, K.-C. (2004). Using amphiphilic pseudo amino acid composition to predict enzyme subfamily classes. *Bioinformatics, 21*(1), 10-19.
2. Peabody, J. D. (2015). Maribavir isomers, compositions, methods of making and methods of using: Google Patents.
3. Wu, R., & Racker, E. (1959). Regulatory mechanisms in carbohydrate metabolism III. Limiting factors in glycolysis of ascites tumor cells. *Journal of Biological Chemistry, 234*(5), 1029-1035.
4. Cai, Y.-D., & Chou, K.-C. (2005). Predicting enzyme subclass by functional domain composition and pseudo amino acid composition. *Journal of proteome research, 4*(3), 967-971.
5. Mitsukura, K., Yoshida, T., & Nagasawa, T. (2002). Synthesis of (R)-2-phenylpropanoic acid from its racemate through an isomerase-involving reaction by Nocardia diaphanozonaria. *Biotechnology letters, 24*(19), 1615-1621.
6. Fischer, G., Bang, H., & Mech, C. (1983). Determination of enzymatic catalysis for the cis-trans-isomerization of peptide binding in proline-containing peptides. *Biomedica biochimica acta, 43*(10), 1101-1111.
7. Wells, W. W., Xu, D. P., Yang, Y., & Rocque, P. A. (1990). Mammalian thioltransferase (glutaredoxin) and protein disulfide isomerase have dehydroascorbate reductase activity. *Journal of Biological Chemistry, 265*(26), 15361-15364.
8. Barman, T. E. (1969). *Enzyme handbook* (Vol. 2): Springer.
9. Walsh, C., Zydowsky, L., & McKeon, F. D. (1992). Cyclosporin A, the cyclophilin class of peptidylprolyl isomerases, and blockade of T cell signal transduction. *Journal of Biological Chemistry, 267*(19), 13115-13118.
10. Torchia, D. A., Sparks, S. W., Young, P. E., & Bax, A. (1989). Proline assignments and identification of the cis K116/P117 peptide bond in liganded staphylococcal nuclease using isotope edited 2D NMR spectroscopy. *Journal of the American Chemical Society, 111*(21), 8315-8317.
11. Feng, Y., Liu, D., & Wang, J. (2003). Native-like partially folded conformations and folding process revealed in the N-terminal large fragments of staphylococcal nuclease: a study by NMR spectroscopy. *Journal of molecular biology, 330*(4), 821-837.
12. Snyder, S. H., & Sabatini, D. M. (1995). Immunophilins and the nervous system. *Nature medicine, 1*(1), 32-37.
13. López-Erauskin, J., Galino, J., Bianchi, P., Fourcade, S., Andreu, A. L., Ferrer, I., . . . Pujol, A. (2012). Oxidative stress modulates mitochondrial failure and cyclophilin D function in X-linked adrenoleukodystrophy. *Brain, 135*(12), 3584-3598.
14. Seizer, P., Gawaz, M., & May, A. E. (2014). Cyclophilin A and EMMPRIN (CD147) in cardiovascular diseases. *Cardiovascular research*, cvu035.
15. Stamnes, M. A., Rutherford, S. L., & Zuker, C. S. (1992). Cyclophilins: a new family of proteins involved in intracellular folding. *Trends in cell biology, 2*(9), 272-276.
16. Stocki, P., Chapman, D. C., Beach, L. A., & Williams, D. B. (2014). Depletion of cyclophilins B and C leads to dysregulation of endoplasmic reticulum redox homeostasis. *Journal of Biological Chemistry, 289*(33), 23086-23096.
17. Elrod, J. W., & Molkentin, J. D. (2013). Physiologic functions of cyclophilin D and the mitochondrial permeability transition pore. *Circulation Journal, 77*(5), 1111-1122.

18. Ying, Y., & Padanilam, B. J. (2016). Regulation of necrotic cell death: p53, PARP1 and cyclophilin D-overlapping pathways of regulated necrosis? *Cellular and Molecular Life Sciences, 73*(11-12), 2309-2324.
19. Baines, C. P., Kaiser, R. A., Purcell, N. H., Blair, N. S., Osinska, H., Hambleton, M. A., . . . Dorn, G. W. (2005). Loss of cyclophilin D reveals a critical role for mitochondrial permeability transition in cell death. *Nature, 434*(7033), 658-662.
20. Mi, H., Kops, O., Zimmermann, E., Jäschke, A., & Tropschug, M. (1996). A nuclear RNA-binding cyclophilin in human T cells. *FEBS letters, 398*(2-3), 201-205.
21. Chothia, C., & Lesk, A. M. (1986). The relation between the divergence of sequence and structure in proteins. *The EMBO journal, 5*(4), 823.
22. Addou, S., Rentzsch, R., Lee, D., & Orengo, C. A. (2009). Domain-based and family-specific sequence identity thresholds increase the levels of reliable protein function transfer. *Journal of molecular biology, 387*(2), 416-430.
23. Doshi, U., Holliday, M. J., Eisenmesser, E. Z., & Hamelberg, D. (2016). Dynamical network of residue–residue contacts reveals coupled allosteric effects in recognition, catalysis, and mutation. *Proceedings of the National Academy of Sciences, 113*(17), 4735-4740.
24. Haile, J. (1992). *Molecular dynamics simulation* (Vol. 18): Wiley, New York.
25. Sugita, Y., & Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters, 314*(1), 141-151.
26. Sagui, C., & Darden, T. A. (1999). Molecular dynamics simulations of biomolecules: long-range electrostatic effects. *Annual review of biophysics and biomolecular structure, 28*(1), 155-179.
27. Goodey, N. M., & Benkovic, S. J. (2008). Allosteric regulation and catalysis emerge via a common route. *Nature chemical biology, 4*(8), 474-482.
28. Case, D. A., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D., . . . Gohlke, H. (2014). Amber 14.
29. Schlick, T., Barth, E., & Mandziuk, M. (1997). Biomolecular dynamics at long timesteps: Bridging the timescale gap between simulation and experimentation. *Annual review of biophysics and biomolecular structure, 26*(1), 181-222.
30. Ryckaert, J.-P., Ciccotti, G., & Berendsen, H. J. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics, 23*(3), 327-341.
31. Verlet, L. (1967). Computer" experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Physical review, 159*(1), 98.
32. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Söding, J. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular systems biology, 7*(1), 539.
33. Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., . . . Lopez, R. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic acids research, 43*(W1), W580-W584.
34. McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., . . . Lopez, R. (2013). Analysis tool web services from the EMBL-EBI. *Nucleic acids research, 41*(W1), W597-W600.

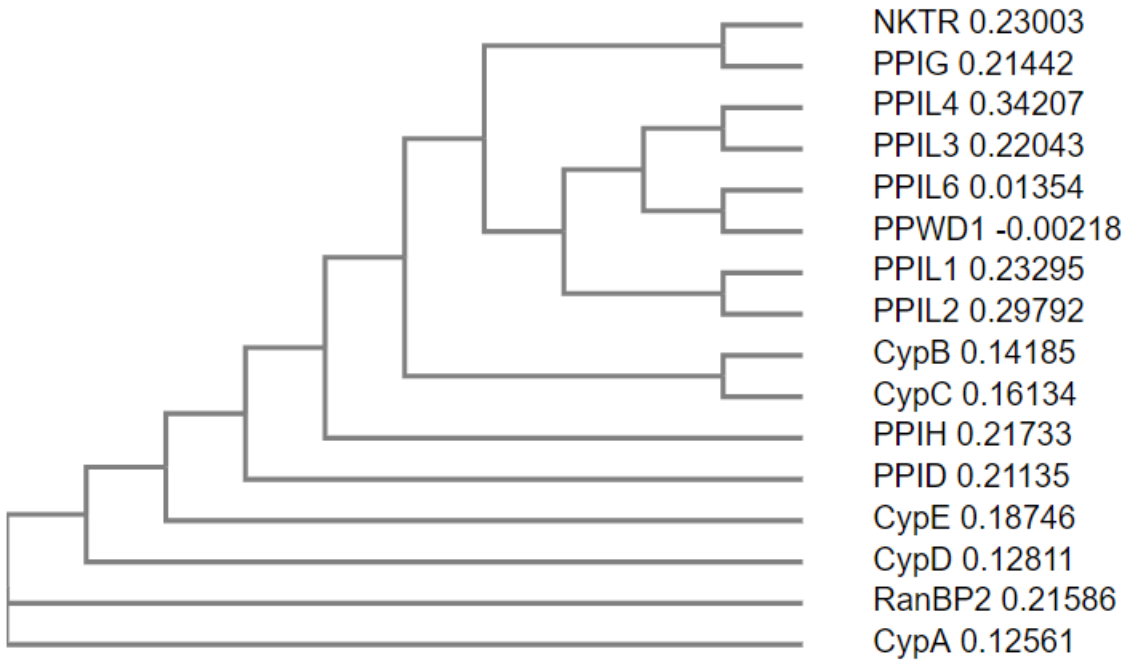
35. Anandakrishnan, R., Aguilar, B., & Onufriev, A. V. (2012). H++ 3.0: automating p K prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic acids research, 40*(W1), W537-W541.
36. Myers, J., Grothaus, G., Narayanan, S., & Onufriev, A. (2006). A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins: Structure, Function, and Bioinformatics, 63*(4), 928-938.
37. Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., & Onufriev, A. (2005). H++: a server for estimating p K as and adding missing hydrogens to macromolecules. *Nucleic acids research, 33*(suppl_2), W368-W371.
38. Bashford, D., & Karplus, M. (1990). pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry, 29*(44), 10219-10225.
39. DeLano, W. (2010). The PyMOL Molecular Graphics System. San Carlos, CA: DeLano Scientific; 2002: Accessed 6/25/2007. Available at http://www. pymol, org.
40. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., . . . Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society, 117*(19), 5179-5197.
41. Doshi, U., & Hamelberg, D. (2009). Reoptimization of the AMBER force field parameters for peptide bond (Omega) torsions using accelerated molecular dynamics. *The Journal of Physical Chemistry B, 113*(52), 16590-16595.
42. Venkatachalam, C., Price, B., & Krimm, S. (1975). A theoretical estimate of the energy barriers between stable conformations of the proline dimer. *Biopolymers, 14*(6), 1121-1132.
43. Karanicolas, J., & Brooks, C. L. (2003). The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: lessons for protein design? *Proceedings of the National Academy of Sciences, 100*(7), 3954-3959.
44. Shehu, A., Kavraki, L. E., & Clementi, C. (2009). Multiscale characterization of protein conformational ensembles. *Proteins: Structure, Function, and Bioinformatics, 76*(4), 837-851.
45. Noel, J. K., Whitford, P. C., & Onuchic, J. N. (2012). The shadow map: a general contact definition for capturing the dynamics of biomolecular folding and function. *The Journal of Physical Chemistry B, 116*(29), 8692-8702.
46. Brinda, K., & Vishveshwara, S. (2005). A network representation of protein structures: implications for protein stability. *Biophysical journal, 89*(6), 4159-4170.
47. Sethi, A., Tian, J., Derdeyn, C. A., Korber, B., & Gnanakaran, S. (2013). A mechanistic understanding of allosteric immune escape pathways in the HIV-1 envelope glycoprotein. *PLoS computational biology, 9*(5), e1003046.
48. Guide, M. U. s. (1998). The mathworks. *Inc., Natick, MA, 5*, 333.
49. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research, 25*(17), 3389-3402.
50. Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schäffer, A. A., & Yu, Y. K. (2005). Protein database searches using compositionally adjusted substitution matrices. *The FEBS journal, 272*(20), 5101-5109.

APPENDICES

Multiple Sequence Alignment of Cyclophilin Isoforms and other PPIase



Multiple Sequence Alignment of Cyclophilin Isoforms and other PPIase.
The PDB FASTA sequences of CypA (1M9F), CypB (3ICH), CypC (2ESL), CypD (4O8H), and CypE (3UCH) and the GenBank FASTA sequences of PPID (CAG46878.1), PPIG (AAH01555.1), PPIH (CAG46697.1), PPIL1 (AAQ89391.1), PPIL2 (CAG30434.1), PPIL3 (NP_570981.1), PPIL4 (Q8WUA2.1), PPIL6 (NP_001265856.1), NKTR (NP_005376.2), RanBP2 (BAA07662.1), and PPWD1 (2A2N_C) were aligned in Clustal Omega.



Phylogenetic Tree of Cyclophilin Isoforms and other PPIase. The phylogenetic tree is represented as a cladogram with actual branch lengths not accurately represented. The branch length values are shown to the right of each protein names.